2016

EMPIRICAL LIKELIHOOD AND DIFFERENTIABLE FUNCTIONALS

Zhiyuan Shen
University of Kentucky, alanshenpku10@gmail.com
Digital Object Identifier: http://dx.doi.org/10.13023/ETD.2016.012

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Zhiyuan Shen, Student

Dr. Mai Zhou, Major Professor

Dr. Constance Wood, Director of Graduate Studies

EMPIRICAL LIKELIHOOD AND DIFFERENTIABLE FUNCTIONALS

DISSERTATION

A dissertation submitted in partial
fulfillment of the requirements for
the degree of Doctor of Philosophy
in the College of Arts and Sciences
at the University of Kentucky

By
Zhiyuan Shen
Lexington, Kentucky

Director: Dr. Mai Zhou, Professor of Statistics
Lexington, Kentucky 2016

ABSTRACT OF DISSERTATION

EMPIRICAL LIKELIHOOD AND DIFFERENTIABLE FUNCTIONALS

Empirical likelihood (EL) is a recently developed nonparametric method of statistical inference. It has been shown by Owen (1988,1990) and many others that empirical likelihood ratio (ELR) method can be used to produce nice confidence intervals or regions. Owen (1988) shows that $-2 \log ELR$ converges to a chi-square distribution with one degree of freedom subject to a linear statistical functional in terms of distribution functions. However, a generalization of Owen's result to the right censored data setting is difficult since no explicit maximization can be obtained under constraint in terms of distribution functions. Pan and Zhou (2002), instead, study the EL with right censored data using a linear statistical functional constraint in terms of cumulative hazard functions. In this dissertation, we extend Owen's (1988) and Pan and Zhou's (2002) results subject to non-linear but Hadamard differentiable statistical functional constraints. In this purpose, a study of differentiable functional with respect to hazard functions is done. We also generalize our results to two sample problems. Stochastic process and martingale theories will be applied to prove the theorems. The confidence intervals based on EL method are compared with other available methods. Real data analysis and simulations are used to illustrate our proposed theorem with an application to the Gini's absolute mean difference.

KEYWORDS: Empirical Likelihood, Statistical Functional, Hadamard Differentiable, Nelson-Aalen Estimator, Survival Analysis, Gini index, Counting Process, Martingale.

Author's signature: Zhiyuan Shen

Date: February 3, 2016

EMPIRICAL LIKELIHOOD AND DIFFERENTIABLE FUNCTIONALS

By
Zhiyuan Shen

Director of Dissertation: Mai Zhou

Director of Graduate Studies: Constance Wood

Date: February 3, 2016

To my wife Xiaowen, and my parents.

ACKNOWLEDGMENTS

First of all, I would like to express my sincerely gratitude to my advisor, Dr. Mai Zhou, for introducing me to the empirical likelihood. His guidance, patience and support made this dissertation possible. Being his student is really a wonderful journey of my Ph.D. studies.

Secondly, I would like to thank my dissertation committee members, Dr. Charnigo, Dr. Chen, Dr. Griffith and Dr. Zheng, for their valuable suggestions and comments.

Finally, I want to thank my wife, Xiaowen Hu. Her support, encouragement and unconditional love help me go through some difficult times in the past three years. I cannot imagine what my life would be without her.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Outline of the Dissertation

This dissertation is organized as follows.

In chapter 2, we review the empirical likelihood ratio tests for both uncensored and right censored data and introduce the Kaplan-Meier (KM) estimator and Nelson-Aalen (NA) estimator. By studying the asymptotic properties of the KM estimator and the NA estimator, we point out that it is more convenient to analyze the right censored data using the hazard functions than using the distribution functions. In this chapter, we also introduce the statistical functional and three distinct derivatives of statistical functionals, Frechet derivative, Hadamard derivative and Gateaux derivative. And we shall focus on the differentiability of the statistical functional in terms of the cumulative hazard functions in later chapters.

In chapter 3, we investigate the Hadamard differentiability of the statistical functional and generalize Pan and Zhou's (2002) results subject to a nonlinear statistical functional in terms of cumulative hazard functions with right censored data.

In chapter 4, again using Hadamard differentiability, we extend Owen's (1988) setting subject to a nonlinear statistical functional in terms of distribution functions for uncensored data.

In chapter 5, we generalize our results in chapter 3 to the two sample problems.

In chapter 6, we compare the confidence intervals based on EL method with other available methods using simulation. QQ plots are used to illustrate our proposed theorem. In particular, we study the Gini's absolute mean difference estimation in detail.

In chapter 7 we discuss the future work to do.

The major contribution of this dissertation is Theorem 3.2.7 in Chapter 3, Theorem 4.2.7 in Chapter 4, Theorem 5.3.6 in Chapter 5 and an application to the Gini's absolute mean difference in Chapter 6.

Chapter 2 Introduction

In this chapter, we briefly review the empirical likelihood ratio tests both with uncensored and right censored data, the Kaplan-Meier and Nelson-Aalen estimators and the statistical functional and its three distinct derivatives. A review of these well-known results sets the stage for later chapters.

2.1 Empirical Likelihood Ratio Test

Empirical Likelihood Ratio Test with Uncensored Data

To facilitate the better understanding of the empirical likelihood ratio test, let's start with parametric likelihood ratio test (PLRT) first. Suppose $X_1, \ldots, X_n$ are n i.i.d. random variables from a population with probability density function (pdf) or probability mass function (pmf) $f(x|\theta_1, \ldots, \theta_p)$, where $\theta_1, \ldots, \theta_p$ are parameters. $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ is a realization of $X_1, \ldots, X_n$. The likelihood function considered as a function of parameters $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_p\}$ is defined by

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = L(\theta_1, \ldots, \theta_p|x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i|\theta_1, \ldots, \theta_p) \tag{2.1}$$

Let Θ denote the full parameter space. The likelihood ratio test is defined as follows.

Definition The likelihood ratio test statistic for testing $H_0 : \theta \in \Theta_0$ *versus* $H_1 : \theta \in \Theta_0^c$ is

$$\lambda(\boldsymbol{x}) = \frac{\sup_{\Theta_0} L(\boldsymbol{\theta}|\boldsymbol{x})}{\sup_{\Theta} L(\boldsymbol{\theta}|\boldsymbol{x})} \tag{2.2}$$

where Θ_0 is some subset of Θ and Θ_0^c is its complement.

The rejection region of likelihood ratio test is of the form $\{\boldsymbol{x} : \lambda(\boldsymbol{x}) \leq c\}$, where c is any number satisfying $0 \leq c \leq 1$.

Wilks (1938) shows that under some regularity conditions, test statistic $-2\log\lambda(\boldsymbol{x})$ is asymptotically $\chi^2_{(p)}$ when the null hypothesis $H_0 : \theta \in \Theta_0$ is true, where p is the number of restrictions imposed on the parameters. This is standard in the textbooks e.g. *Casella and Berger (1990, Chapter 10).*

As we have the Wilks theorem, the PLRT can be used to test hypothesis and generate confidence intervals and regions. However, it is applicable only when we know what distribution or density $f(x|\boldsymbol{\theta})$ the random variables are from. In some cases, the parametrical distribution or density we assume is questionable. If this is the case, empirical likelihood ratio test (ELRT) may be used, which does not require strong distribution or density assumptions.

The ELRT is a recently developed nonparametric method of statistical inference. It has been shown by Owen (1988,1990) and many others that empirical likelihood ratio (ELR) method can be used to produce nice confidence intervals or regions in ways that are analogous to those of PLRT , but without strong distribution or density assumptions.

Suppose $X_1, \ldots, X_n$ are n i.i.d. random variables with an unknown distribution function F_0 and $x_1, \ldots, x_n$ is a realization of $X_1, \ldots, X_n$. Owen (1988) defined the empirical likelihood function in terms of distribution functions as follows.

$$EL(F) = \prod_{i=1}^{n} \Delta F(x_i) = \prod_{i=1}^{n} p_i \tag{2.3}$$

where $p_i = \Delta F(x_i) = F(x_i) - F(x_i-)$ and $F(t-)$ is the left continuous version of $F(t)$.

It can be shown that empirical distribution function $\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} I[X_i \leq t]$ maximizes $EL(F)$ (2.3) among all possible distribution functions, which is well-known as nonparametric maximum likelihood estimator (NPMLE) of F_0.

Owen (1988) proves the nonparametric version of Wilks's theorem. He defines the empirical likelihood ratio (ELR) function as follows.

$$ELR = \frac{EL(F)}{EL(\hat{F}_n)} \tag{2.4}$$

where $EL(\cdot)$ is defined in (2.3) and $\hat{F}_n$ is the empirical distribution function.

He shows that $-2\log ELR(\theta_0)$ converges to a $\chi^2_{(1)}$ when the null hypothesis H_0 : $\int g(t)dF(t) = \theta_0$ is true, where $ELR(\theta_0)$ is the maximum of the empirical likelihood ratio function (2.4) subject to a linear functional constraint of F, $\int g(t)dF(t) = \theta_0$ and $\theta_0 = \int g(t)dF_0(t)$.

Empirical Likelihood Ratio Test with Right Censored Data

Suppose $X_1, \ldots, X_n$ are n i.i.d. random variables with distribution function F_0 denoting lifetimes and $C_1, \ldots, C_n$ are n i.i.d. random variables with distribution function G_0 denoting censoring times. C is independent of X. Only censored observations are available to us.

$$T_i = \min(X_i, C_i), \quad \delta_i = I[X_i \leq C_i], \quad i = 1, \ldots, n \tag{2.5}$$

The empirical likelihood function in terms of distribution functions based on the censored observations pertaining F is

$$EL(F) = \prod_{i=1}^{n} [\Delta F(T_i)]^{\delta_i} [1 - F(T_i)]^{1-\delta_i} \tag{2.6}$$

where $\Delta F(T_i) = F(T_i) - F(T_i-),\ i = 1, \ldots, n$. See *Owen (2001), Empirical Likelihood, Chapter 6* for a discussion of the above empirical likelihood function.

To generalize Owen's setting to the right censored data, we need to maximize $EL(F)$ (2.6) both without and without the constraint $\int g(t)dF(t) = \theta_0$, where $\theta_0 = \int g(t)dF_0(t)$. As is well known, the Kaplan-Meier estimator maximizes $EL(F)$ without the constraint (Kaplan and Meier (1958)). We will introduce the Kaplan-Meier estimator in the next section. To maximize the $EL(F)$ (2.6) under the linear constraint $\int g(t)dF(t) = \theta_0$, we utilize the Lagrange multiplier method. Denote $w_i = \Delta F(T_i)$ and notice that $\sum_{i=1}^n w_i = 1$, since the summation of all jumps of a discrete distribution function equals to one, we can write the constraint $\int g(t)dF(t) = \theta_0$ in the discrete format as follows.

$$\sum_{i=1}^n \delta_i g(T_i) w_i = \theta_0 \tag{2.7}$$

The $\log EL(F)$ in terms of w_i is as follows.

$$\log EL(F) = \sum_{i=1}^n \left(\delta_i \log w_i + (1-\delta_i) \log \left(1 - \sum_{j=1}^i w_j \right) \right) \tag{2.8}$$

In order to apply the Lagrange multiplier method, we form the target function G as follows.

$$\begin{aligned} G = & \sum_{i=1}^n \left[\delta_i \log w_i + (1-\delta_i) \log \left(1 - \sum_{j=1}^i w_j \right) \right] \\ & + \gamma \left(1 - \sum_{i=1}^n w_i \right) + n\lambda \left(\theta_0 - \sum_{i=1}^n \delta_i g(T_i) w_i \right) \end{aligned} \tag{2.9}$$

Taking the derivative with respect to $w_i,\ i = 1, \ldots, n$ and equaling them to 0 yields

$$\frac{\partial G}{\partial w_i} = \frac{\delta_i}{w_i} - \sum_{l=i}^n (1-\delta_l) \frac{1}{1 - \sum_{j=1}^l w_j} - \gamma - n\lambda \delta_i g(T_i) = 0 \tag{2.10}$$

Then we have

$$\gamma = \frac{\delta_i}{w_i} - \sum_{l=i}^{n} (1-\delta_l) \frac{1}{1-\sum_{j=1}^{l} w_j} - n\lambda\delta_i g(T_i) \tag{2.11}$$

Multiplying w_i on both sides and taking the summation through 1 to n gives us

$$\begin{aligned} \gamma = \sum_{i=1}^{n} w_i\gamma &= \sum_{i=1}^{n} \delta_i - \sum_{i=1}^{n}\sum_{l=i}^{n} \left[(1-\delta_l) \frac{1}{1-\sum_{j=1}^{l} w_j}\right] w_i - n\lambda \sum_{i=1}^{n} \delta_i g(T_i) w_i \\ &= \sum_{i=1}^{n} \delta_i - \sum_{l=1}^{n} (1-\delta_l) \frac{\sum_{j=1}^{l} w_j}{1-\sum_{j=1}^{l} w_l} - n\lambda\theta_0 \end{aligned} \tag{2.12}$$

since $\sum_{i=1}^{n} \delta_i g(T_i) w_i = \theta_0$ and $\sum_{i=1}^{n} w_i = 1$.

Plugging γ into (2.10), we have an equation for w_i

$$\frac{\delta_i}{w_i} = (n-i+1) + \sum_{l=1}^{i-1} \delta_l - \sum_{l=1}^{i-1} (1-\delta_l) \frac{\sum_{j=1}^{l} w_l}{1-\sum_{j=1}^{l} w_l} - n\lambda\theta_0 + n\lambda\delta_i g(T_i) \tag{2.13}$$

From the above equation, we can see that w_i is a non-linear function of all of its previous jumps $w_j,\ j = 1, \ldots, i-1$. (2.13) is a recursive formula for computing w_i. Although we can solve the problem computationally (Zhou and Yang (2015)), it is difficult to solve it analytically, since no explicit maximization can be obtained under the constraint in terms of distribution functions.

Pan and Zhou (2002) generalize Owen's result to the right censored data by using a linear functional constraint in terms of the cumulative hazard function. The relationship between the distribution function and hazard function is as follows.

$$1 - F(t) = \prod_{s \le t} (1 - \Delta\Lambda(s)) \qquad \Delta\Lambda(t) = \frac{\Delta F(t)}{1 - F(t-)} \tag{2.14}$$

The empirical likelihood function (2.6) can be rewritten in terms of hazard functions as follows.

$$EL(\Lambda) = \prod_{i=1}^{n} \left\{ [\Delta\Lambda(T_i)]^{\delta_i} \left[\prod_{j:T_j<T_i} (1-\Delta\Lambda(T_j)) \right]^{\delta_i} \left[\prod_{j:T_j\leq T_i} (1-\Delta\Lambda(T_j)) \right]^{1-\delta_i} \right\} \tag{2.15}$$

A simpler version can be obtained if we merge the second and third terms of the above equation and replace it with $\exp\{-\Lambda(T_i)\}$, which is called the Poisson extension of the likelihood introduced by Murphy (1995).

$$AL(\Lambda) = \prod_{i=1}^{n} [\Delta\Lambda(T_i)]^{\delta_i} \exp\{-\Lambda(T_i)\} \tag{2.16}$$

For a detailed discussion of different extensions of the likelihood function for discrete distributions, see Gill (1989). See Pan and Zhou (2002) for a discussion of the legitimacy of the use of AL.

Pan and Zhou (2002), study EL with right censored data using a linear functional constraint in terms of the cumulative hazard functions. They define the empirical likelihood ratio (ELR) function as follows.

$$ELR = \frac{EL(\Lambda)}{EL(\hat{\Lambda}_{NA})} \tag{2.17}$$

where $EL(\Lambda)$ is defined in (2.15) and $\hat{\Lambda}_{NA}$ is the so-called Nelson-Aalen estimator which maximizes $EL(\Lambda)$ (2.15) among all cumulative hazard functions (Nelson (1969, 1974), Aalen (1976)). We will introduce the Nelson-Aalen estimator in the next section.

Pan and Zhou (2002) prove that $-2\log ELR(\theta_0)$ has an asymptotic $\chi^2_{(1)}$ when the null hypothesis $H_0 : \int g(t) d\Lambda(t) = \theta_0$ is true, where $ELR(\theta_0)$ is the maximum of the ELR function (2.17) subject to a linear functional constraint of Λ, $\int g(t) d\Lambda(t) = \theta_0$ where $\theta_0 = \int g(t) d\Lambda_0(t)$ and Λ_0 is cumulative hazard function associated with F_0 defined in (2.5).

2.2 Kaplan-Meier Estimator and Nelson-Aalen Estimator

Kaplan-Meier Estimator

The Kaplan-Meier (KM) estimator, also known as the product-limit estimator, is a nonparametric statistic to estimate the survival probability with lifetime data. It was first introduced by Kaplan and Meier in 1958. The survival function represents the probability that a subject from a given population has a lifetime exceeding time t. Let $X_1, \ldots, X_n$ be n i.i.d. random variables with survival function $S(t) = P(X > t)$ denoting lifetimes and $C_1, \ldots, C_n$ be n i.i.d. random variables with survival function $G(t) = P(C > t)$ denoting censoring times. X and C are independent. And only censored observations are available to us.

$$T_i = \min(X_i, C_i), \quad \delta_i = I(X_i \leq C_i), \quad i = 1, \ldots, n \tag{2.18}$$

Suppose there are k distinct uncensored lifetimes $t_1 < t_2 < \ldots < t_k$. Corresponding to each t_i we have n_i, the number of individuals at risk prior to time t_i and d_i, the number of deaths at time t_i. The KM estimator is defined as follows.

$$\hat{S}_{KM}(t) = \prod_{t_i \leq t} \frac{n_i - d_i}{n_i} \tag{2.19}$$

Kaplan and Meier shows that $\hat{S}_{KM}$ is a nonparametric maximum likelihood estimator of survival function $S(t)$ in the sense that it maximizes the following likelihood function

$$L(S) = \prod_{uncensored} [S(Z_i-) - S(Z_i)] \prod_{censored} S(Z_i) \tag{2.20}$$

over the parameter space $\Theta = \{all\ survival\ functions\}$. The variance estimator of the KM estimator, which is the well-known Greenwood formula, is defined as follows.

$$\hat{Var}\left(\hat{S}(t)\right) = \hat{S}_{KM}(t)^2 \sum_{t_i \le t} \frac{d_i}{n_i\left(n_i - d_i\right)} \tag{2.21}$$

Breslow and Crowley (1974) first derive the asymptotic properties of Kaplan-Meier estimator under a random censorship model. They show that $\hat{S}_{KM}(t)$ is asymptotically normal and its asymptotic variance can be estimated by Greenwood formula consistently.

More recent references of asymptotic results of the KM estimator utilize counting process and martingale theories. We will discuss them with the Nelson-Aalen estimator.

Nelson-Aalen Estimator

The Nelson-Aalen (NA) estimator was first proposed by Nelson (1969,1972). Its asymptotic properties were studied by Breslow and Crowley (1974) and by Aalen (1976).

While the Kaplan-Meier (KM) estimator is the nonparametric maximum likelihood estimator (NPMLE) of the survival functions, the Nelson-Aalen (NA) estimator is the NPMLE of the cumulative hazard functions. It is defined as

$$\hat{\Lambda}_{NA}(t) = \sum_{t_i \le t} \frac{d_i}{n_i} \tag{2.22}$$

where d_i is the number of deaths at time t_i and n_i is the number of individuals at risk prior to time t_i.

It is much more mathematically convenient to use cumulative hazard functions instead of distribution functions to analyze the right censored data, because the NA estimator has a lot of properties that the KM estimator does not have.

First, the NA estimator can be represented in a form of martingale while the KM estimator can also be represented in a form of martingale but in a complex format. Second, with the knowledge of counting process and martingale theory, we learn that the predictable integration with respect to a martingale is also a martingale. This implies that the predictable integration with respect to the NA estimator is also a martingale. To represent the NA estimator and the predictable integration with respect to the NA estimator in the form of martingales is important, since **Martingale Central Limit Theorem** can be applied to obtain their asymptotic properties. See *Kalbfleisch and Prentice (2002), Chapter 5* for more details about the Martingale Central Limit Theorem.

It is helpful to define two technical terms before we introduce the definition of the martingale **(Kalbfleisch and Prentice (2002))**.

1. A stochastic process $U = \{U(t), t \geq 0\}$ is said to be adapted to the filtration $\mathscr{F}_t$, if for each t, $U(t)$ is a function of (or is specified by) $\mathscr{F}_t$. In measure-theoretic terms, U is said to be adapted if $U(t)$ is $\mathscr{F}_t$ measurable for each $t \in [0, \tau]$. In less formal terms, this simply means that the value of $U(t)$ is fixed once $\mathscr{F}_t$ is given.

2. The stochastic process U is said to be predictable with respect to the filtration $\mathscr{F}_t$, if for each t, the value of $U(t)$ is a function of (or is specified by) $\mathscr{F}_{t-}$. Again, in measure-theoretic terms, U is predictable if $U(t)$ is $\mathscr{F}_{t-}$ measurable for all $t \in [0, \tau]$.

For example, if $f(t)$ is left continuous with respect to t, $f(t)$ is predictable.

The definition of the martingale cited from Kalbfleisch and Prentice (2002) is as follows.

Definition (Kalbfleisch and Prentice (2002)) A (real-valued) stochastic process $\{M(t),\ 0 \leq t \leq \tau\}$ is a martingale with respect to the filtration $\{\mathscr{F}_t\}$ if it is cadlag, adapted to $\mathscr{F}_t$, and satisfies the martingale property

$$E[M(t)|\mathscr{F}_t] = M(s) \quad for\ all\ s \leq t \leq \tau \tag{2.23}$$

or equivalently

$$E[dM(t)|\mathscr{F}_{t-}] = 0 \quad for\ all\ t \in (0, \tau] \tag{2.24}$$

For example, suppose X is a continuous random variable with hazard function $h(t)$, we define the one jump counting process as $N(t) = I[X \leq t]$. It can be shown that

$$M(t) = N(t) - \int_0^t h(s)I[X \geq s]ds \tag{2.25}$$

is a martingale with respect to $\mathscr{F}_t$.

Next, we introduce the **predictable variation process**. The primary role of studying the predictable variation process is to compute the variance of the counting process martingale $M(t)$ and the variance of the integration with respect to $M(t)$. Also the conditions of the **Martingale Central Limit Theorem** are formulated in terms of the predictable variation process.

Definition (Kalbfleisch and Prentice (2002)) The predictable variation process of a square integrable martingale M is defined as

$$\langle M \rangle(t) = \int_0^t var[dM(u)|\mathscr{F}_{u-}] \tag{2.26}$$

Equivalently, we can write

$$d\langle M \rangle(t) = var[dM(t)|\mathscr{F}_{t-}] \tag{2.27}$$

The martingale $M(t)$ is said to be square integrable if $E[M^2(\tau)] < \infty$.

For example, the predictable variation process of the one jump counting process martingale $M(t) = I[X \le t] - \int_0^t h(s)I[X \ge s]ds$ is

$$\langle M\rangle(t) = \langle I[X \le t] - \int_0^t h(s)I[X \ge s]ds\rangle = \int_0^t h(s)I[X \ge s]ds \tag{2.28}$$

where X is a random variable with continuous hazard function $h(t)$.
Another important example is the predictable variation process of the integration with respect to a martingale $M(t)$.

$$\langle \int_0^t f(s)dM(s)\rangle = \int_0^t f^2(s)d\langle M(s)\rangle \tag{2.29}$$

where $f(t)$ is predictable.

Preceding the martingale representation of the Nelson-Aalen estimator, we claim an important theorem.

Theorem 2.2.1 (Anderson, P.K. et al. (1993)) *Suppose M is a finite variation local square integrable martingale, H is a predictable process, and $\int H^2 d\langle M\rangle$ is locally finite. Then $\int HdM$ is a local square integrable martingale.*

Let $X_1, \ldots, X_n$ be n i.i.d. random variables with distribution function $F(t)$, cumulative distribution function $\Lambda(t)$ and hazard function $h(t)$ denoting lifetimes and $C_1, \ldots, C_n$ be n i.i.d. random variables with distribution function $G(t)$ denoting censoring times. X and C are independent. And only censored observations are available to us.

$$T_i = \min(X_i, C_i), \quad \delta_i = I(X_i \le C_i), \quad i = 1, \ldots, n \tag{2.30}$$

Define $M_n(t)$ as follows

$$M_n(t) = \sum_{i=1}^{n} M_i(t) = \sum_{i=1}^{n} \left(I[X_i \leq t, \delta_i = 1] - \int_0^t h(s) I[X_i \geq s] ds \right) \tag{2.31}$$

It can be shown that $M_n(t)$ is a local square integrable martingale with respect to $\mathscr{F}_t$ (Anderson et al.(1993)).

Denote $R(t) = \sum_{j=1}^{n} I[T_j \geq t]$. Since $R(t)$ is left continuous, it is predictable. The Nelson-Aalen estimator can be represented as a predictable integration with respect to a local square integrable martingale $M_n(t)$ as follows (Anderson et al. (1993)).

$$\hat{\Lambda}_{NA}(t) - \Lambda(t) = \int_0^t \frac{1}{R(s)} dM_n(s) \tag{2.32}$$

Therefore, by Theorem 2.2.1, (2.32) is also a martingale. It can also been shown that two conditions of the Martingale Central Limit Theorem are satisfied. Then we have

$$\sqrt{n} \left(\hat{\Lambda}_{NA}(t) - \Lambda(t) \right) \xrightarrow{D} BM(A(t)) \tag{2.33}$$

where $BM(t)$ is a standard Brownian motion and

$$A(t) = \int_0^t \frac{d\Lambda(s)}{P(X \geq s)} = \int_0^t \frac{d\Lambda(s)}{(1 - F(s-))(1 - G(s-))} \tag{2.34}$$

For more details about the asymptotic properties of the NA estimator, see Anderson et al. (1993).

The Kaplan-Meier estimator can also be represented as a martingale but in a complex format (Gill (1983)). Suppose $(T_1, \delta_1), \ldots, (T_n, \delta_n)$ are n i.i.d. random vectors defined in (2.30). Let $\hat{F}_{KM}(t)$ be the product-limit estimator such that $1 - \hat{F}_{KM}(t) = \hat{S}_{KM}(t) = \prod_{j:T_j \leq t} \left(1 - \frac{d_j}{n_j}\right)$ and $1 - H = (1 - F)(1 - G)$.

Theorem 2.2.2 (Gill (1983)) *For any τ such that $H(\tau-) < 1$*

$$\sqrt{n}\left(\frac{\hat{F}_{KM} - F}{1 - F}\right) \xrightarrow{D} BM(C) \text{ in } \mathbb{D}[0,\tau] \text{ as } n \ \to \infty \tag{2.35}$$

where $BM(t)$ is a standard Brownian motion, $\mathbb{D}[0,\tau]$ is the set of all cadlag functions and

$$\Lambda(t) = \int_0^t \frac{dF(s)}{1 - F(s-)}, \quad C(t) = \int_0^t \frac{dF(s)}{\left(1 - F(s-)\right)^2 \left(1 - G(s-)\right)} = \int_0^t \frac{d\Lambda(s)}{1 - H(s-)} \tag{2.36}$$

Remark Note that $BM(C)$ is a continuous Gaussian martingale, zero at time zero, with covariance function

$$Cov\left[BM(C(s)), BM(C(t))\right] = C(s) \wedge C(t) = C(s \wedge t) \tag{2.37}$$

where $\wedge$ denotes minimum. ∎

Although we represent the Kaplan-Meier estimator as a martingale, we have $1 - F$ in the denominator. Unfortunately, what we are interested in is the integration with respect to $\hat{F}_{KM} - F$, which is not a martingale, so Theorem 2.2.1 cannot be applied to obtain the asymptotic properties of the integration with respect to $\hat{F}_{KM} - F$. Nevertheless, $\hat{\Lambda}_{NA} - \Lambda$ is a martingale and by Theorem 2.2.1, the integration with respect to $\hat{\Lambda}_{NA} - \Lambda$ is also a martingale, the Martingale Central Limit Theorem can be applied to obtain its asymptotic properties.

Akritas (2000) proves a central limit theorem for the integration with respect to $\hat{F}_{KM} - F$ but not on the whole real line. Suppose $(T_1, \delta_1), \ldots, (T_n, \delta_n)$ are n i.i.d. random vectors defined in (2.30). Let $\hat{F}_{KM}(t)$ be the product-limit estimator s.t. $1 - \hat{F}_{KM}(t) = \hat{S}_{KM}(t) = \prod_{j:T_j \le t}\left(1 - \frac{d_j}{n_j}\right)$ and $1 - H = (1-F)(1-G)$, $S = 1 - F$.

Let $\tau_n = \max(X_1, \ldots, X_n)$ and $\tau_F = \sup\{x : F(x) < 1\}$, for any distribution function F. Let $\phi : \mathbb{R} \to \mathbb{R}$ be any measurable function such that $\int \phi^2 dF < \infty$, where $\mathbb{R}$ is the real line.

Theorem 2.2.3 (Akritas (2000)) *Let the following assumption hold,*

$$\int_{-\infty}^{\tau_H} \frac{\phi^2(s)}{1 - G(s-)} dF(s) < \infty \tag{2.38}$$

Then if $\tau_n < \tau_F$ a.s. or $\phi(\tau_F) = 0$,

$$\sqrt{n} \int_{-\infty}^{\tau_H} \phi(s) d\left(\hat{F}_{KM}(s) - F(s)\right) \xrightarrow{D} N(0, \sigma^2) \tag{2.39}$$

where $\sigma^2 = \int_{-\infty}^{\tau_H} \frac{S(s)}{1-H(s-)} \left[\phi(s) - \bar{\phi}(s)\right]^2 dF(s)$ and $\bar{\phi}(s) = \frac{1}{S(s)} \int_{(s,\tau_H]} \phi(t) dF(t)$.

Although Akritas proves a central limit theorem for the integration with respect to the KM estimator, he approximates it by an integration with respect to NA estimator plus a small error $o_p(\frac{1}{\sqrt{n}})$. This means that he takes the advantage of the convenient martingale expression of the NA estimator to obtain the asymptotic property of the KM estimator. Therefore, Akritas's theorem strengthens our claim that it is more convenient to analyze the right censored data using hazard functions than using distribution functions. For more details, see Akritas (2000).

From the discussion in this section, it is clear that the NA estimator has plenty of nice properties for deriving the asymptotic properties while the KM estimator seldom has. This is the major reason why hazard functions are more frequently used than distribution functions with right censored data. In the next section, we introduce the statistical functional and its three distinct differentiability. We would initially introduce the statistical functional in terms of distribution functions. Since hazard functions are going to be used to analyze the right censored data, eventually, we shall focus on the statistical functional in terms of cumulative hazard functions and their differentiability.

2.3 Statistical Functional and Its Derivative

Statistical Functional

In this dissertation, we investigate the large sample properties of empirical likelihood ratio subject to various kinds of non-linear constraints. A constraint can frequently be formulated via a statistical functional on a normed linear space. And a statistical functional with differentiability properties will provide a handle to work out its asymptotic behavior.

A lot of work has already been done on statistical functional in terms of distribution functions. They were first introduced by von Mises (1936, 1937, 1947). His work was largely ignored until late 1960s when the development of robust statistics had a boom. Since then, von Mises's theory has been studied and extended by several authors in different directions: Filippova (1962), Reeds (1976), Huber (1977,1981) and Serfling (1980). Now von Mises's calculus has been widely used in the theory of robust estimation and study of bootstrap methods.

To have a better understanding of von Mises's method, let's begin with a discussion of empirical distribution function. Let $X_1, \ldots, X_n$ be n i.i.d. random variables with distribution function $F(x)$.

Definition The empirical distribution function $\hat{F}_n(x)$ is the cumulative distribution function that puts mass $\frac{1}{n}$ at each data point X_i.

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I[X_i \leq x] \tag{2.40}$$

By **Strong Law of Large Numbers**, the empirical distribution function $\hat{F}_n(x)$ converges to $F(x)$ almost surely, for every value of $x \in \mathbb{R}$.

$$\hat{F}_n(x) \xrightarrow{a.s.} F(x) \ as \ n \to \infty \tag{2.41}$$

The **Glivenko-Cantelli Theorem** states a stronger result that the convergence in fact happens uniformly over $x \in \mathbb{R}$.

$$\sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F(x) \right| \xrightarrow{a.s.} 0 \ as \ n \to \infty \tag{2.42}$$

By **Central Limit Theorem**, for any fixed $x \in \mathbb{R}$, $\hat{F}_n(x)$ has an asymptotic normal distribution.

$$\sqrt{n}\left(\hat{F}_n(x) - F(x)\right) \xrightarrow{D} N\left(0, F(x)\left(1 - F(x)\right)\right) \ as \ n \to \infty \tag{2.43}$$

The **Donsker's Theorem** extends the above result (2.43) and asserts that the empirical process $\sqrt{n}\left(\hat{F}_n - F\right)$, which is indexed by $x \in \mathbb{R}$, converges in distribution to the mean-zero Gaussian process $G_F = B(F(x))$, where B is the standard Brownian bridge. The covariance structure of the Gaussian process is

$$E\left[G_F(t_1)G_F(t_2)\right] = F(t_1 \wedge t_2) - F(t_1)F(t_2) \tag{2.44}$$

where $\wedge$ denotes the minimum.

In parametric statistics, if we have worked out the asymptotic distribution of $\hat{\theta}$ as $\sqrt{n}\left(\hat{\theta} - \theta_0\right)$ converges to a normal distribution, the delta method can be used to obtain the asymptotic distribution for a function of $\hat{\theta}$, $T(\hat{\theta})$. In nonparametric statistics, since the asymptotic normality of $\hat{F}_n - F$ is well known as we discussed above, we use the **Functional Delta Method** to obtain the asymptotic distribution for $T(\hat{F}_n) - T(F)$, where $T(F)$ is a functional of F.

To facilitate a better understanding, we start with a simpler version of functional delta method assuming that $T(F) = \int a(x)dF(x)$, which is called a linear functional. But ultimately we shall work with the non-linear functional.

First, we introduce the **Influence Curve**, which will be used in establishing the Functional Delta Method later. The influence curve was first introduced by Hampel (1974) and used in robust estimation. It also provides a way to compute the asymptotic variance when the statistic is asymptotically normal. The definition of the influence curve (Hampel (1974)) is as follows.

Definition (Hampel (1974)) Let $\mathbb{R}$ be the real line and T be a real-valued functional defined on some subset of the set of all probability measures on $\mathbb{R}$, and let F denote a probability measure on $\mathbb{R}$ for which T is defined. Denote by δ_x the probability measure determined by the point mass 1 in any given point $x \in \mathbb{R}$. Mixtures of F and some δ_x are written as $(1-\epsilon)F + \epsilon\delta_x$, for $0 < \epsilon < 1$. Then the influence curve $IC_{T,F}(x)$ of T at F is defined pointwisely by

$$IC_{T,F}(x) = \lim_{\epsilon \downarrow 0} \frac{\{T\left[(1-\epsilon)F + \epsilon\delta_x\right] - T(F)\}}{\epsilon} \tag{2.45}$$

if this limit is defined for every point $x \in \mathbb{R}$. ∎

Theorem 2.3.1 (Wasserman (2006)) *Let $T(F) = \int a(x)dF(x)$ be a linear functional and $L_F(x)$ be the influence curve of T at F. Then we have*

1. $L_F(x) = a(x) - T(F)$ *and* $\hat{L}(x) = L_{\hat{F}_n}(x) = a(x) - T(\hat{F}_n)$

2. For any distribution function G,

$$
\begin{aligned}
T(G) &= T(F) + \int L_F(x) dG(x) \\
&= T(F) + \int \left(a(x) - T(F)\right) dG(x)
\end{aligned}
\tag{2.46}
$$

3. $\int L_F(x)dF(x) = 0,$ *where* $L_F(x) = a(x) - T(F)$.

4. Let $\tau^2 = \int L_F^2(x)dF(x)$. *Then if* $\tau^2 < \infty$,

$$
\sqrt{n}\left(T(\hat{F}_n) - T(F)\right) \xrightarrow{D} N(0, \tau^2) \text{ as } n \to \infty \tag{2.47}
$$

where $L_F(x) = a(x) - T(F)$.

5. Let

$$
\hat{\tau}^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{L}^2(X_i) = \frac{1}{n}\sum_{i=1}^{n}\left(a(X_i) - T(\hat{F}_n)\right)^2 \tag{2.48}
$$

Then

$$
\hat{\tau}^2 \xrightarrow{P} \tau^2 \text{ and } \frac{\hat{se}}{se} \xrightarrow{P} 1, \text{ as } n \to \infty \tag{2.49}
$$

where $\hat{se} = \frac{\hat{\tau}}{\sqrt{n}}$ *and* $se = \sqrt{\frac{F(x)(1-F(x))}{n}}$.

6.

$$
\frac{\sqrt{n}\left(T(\hat{F}_n) - T(F)\right)}{\hat{\tau}} \xrightarrow{D} N(0, 1), \text{ as } n \to \infty \tag{2.50}
$$

Proof See *Wasserman, L. (2006), All of Nonparametric Statistics, Chapter 2.* ∎

If the functional $T(F)$ is non-linear, which means not of the form $T(F) = \int a(x)dF(x)$, (2.46) will not hold exactly, but it may hold approximately. We summarize the approach to this problem in the following paragraphs. For more detailed discussion, see *Fernholz (1983),Chapter I.*

To deal with the non-linear functional $T(F)$, von Mises came up with a Taylor expansion of a statistic $T(\hat{F}_n)$ as follows.

$$T(\hat{F}_n) = T(F) + T_F^{'}(\hat{F}_n - F) + Rem(\hat{F}_n - F) \tag{2.51}$$

where $\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I[X_i \leq x]$ is the empirical distribution function defined in (2.40). $T_F^{'}$ is the derivative of T at F. We will introduce three distinct derivatives of T at F later in this section. In particular, the term $T_F^{'}(\hat{F}_n - F)$ is linear and is therefore a sum of i.i.d. random variables, so the central limit theorem implies that for some finite $\sigma^2 > 0$,

$$\sqrt{n}T_F^{'}\left(\hat{F}_n - F\right) \xrightarrow{D} N(0, \sigma^2) \tag{2.52}$$

Under some conditions, the remaining term

$$\sqrt{n}Rem(\hat{F}_n - F) \xrightarrow{P} 0 \tag{2.53}$$

is satisfied. If (2.52) and (2.53) hold, by the Slutsky theorem, we have

$$\sqrt{n}\left(T(\hat{F}_n) - T(F)\right) \xrightarrow{D} N(0, \sigma^2) \tag{2.54}$$

The following is a brief discussion of the conditions that make (2.53) satisfied. For more detailed discussions, see Reeds (1976) and Fernholz (1983).

The satisfaction of (2.53) relies on the differentiability of statistical functional T. Different authors choose different conditions to make (2.53) satisfied. They can be mainly classified into three types: Gateaux differentiability, Hadamard differentiability and Frechet differentiability. Gateaux differentiability is a weak form of differentiability and various authors have to supplement Gateaux differentiability with extra conditions e.g. the second order derivative, which is seldom satisfied. The assumption of Frechet differentiability also implies (2.53), but is still too strong because few statistic is Frechet differentiable. The most popular condition that implies (2.53) is the Hadamard differentiability, which is a weaker form of differentiability than Frechet differentiability and applicable to a large class of statistics.

Next, we introduce the statistical functional and its three distinct differentiability, Frechet differentiability, Hadamard differentiability and Gateaux differentiability. From the above discussion, we shall focus on Hadamard differentiability in later chapters.

Now let's continue with the definition of statistical functional in terms of distribution functions (Fernholz (1983)).

Definition (Fernholz (1983)) Let $X_1, \ldots, X_n$ be a random sample from distribution function F and let $T_n = T_n(X_1, \ldots, X_n)$ be a statistic. If T_n can be written as a functional T of the empirical distribution function $\hat{F}_n$, $T_n = T(\hat{F}_n)$, where $\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I[X_i \leq x]$ and T does not depend on n, then T is called a statistical functional. The domain of definition of T is assumed to contain the empirical distribution function $\hat{F}_n$ for all $n \geq 1$, as well as the population distribution function F. Unless otherwise specified, the range of T will be the set of real numbers. The parameter to be estimated is $T(F)$. ▮

The following is a simple example of statistical functional in terms of distribution functions.

Let ϕ be a real valued function and $\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} I[X_i \leq t]$ be the empirical distribution function and let

$$T_n(X_1,\ldots,X_n) = \frac{1}{n}\sum_{i=1}^{n}\phi(X_i) = \int \phi(t)d\hat{F}_n(t) \tag{2.55}$$

Then for a general distribution function G, the functional defined by

$$T(G) = \int \phi(x)dG(x) \tag{2.56}$$

satisfies $T_n(X_1,\ldots,X_n) = T(\hat{F}_n)$. And functionals of this form are called linear statistical functionals in terms of distribution functions.

Any statistical functional in terms of distribution functions not of the form $T(G) = \int \phi(x)dG(x)$ is called a non-linear statistical functional in terms of distribution functions.

For example, of a non-linear function, let

$$T_n(X_1,\ldots,X_n) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \tag{2.57}$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. For a general distribution function G, the functional is defined by

$$T(G) = \frac{1}{2}\int\int (x-y)^2\, dG(x)dG(y) \tag{2.58}$$

Then

$$
\begin{aligned}
T(\hat{F}_n) &= \frac{1}{2}\int\int (x-y)^2\, d\hat{F}_n(x)d\hat{F}_n(y) \\
&= \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - X_j)^2 \\
&= \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(X_i^2 - 2X_iX_j + X_j^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \left(\bar{X}\right)^2 = T_n(X_1,\ldots,X_n)
\end{aligned}
\tag{2.59}
$$

T is a statistical functional but not of the form $\int \phi(x)dG(x)$, which is called a non-linear statistical functional in terms of distribution functions.

In the previous section, we learn that the Kaplan-Meier estimator and the Nelson-Aalen estimator are nonparametric maximum likelihood estimator (NPMLE) for the survival functions and the cumulative hazard functions respectively. Moreover, the NA estimator can be expressed in a form of martingale while the KM estimator can not. And it is more convenient to use the hazard functions instead of the distribution functions for the right censored data. In order to use the hazard functions, we shall write the empirical likelihood in terms of them, which we already did in (2.15). To compute the empirical likelihood ratio, we need to calculate the maximum of the empirical likelihood under a statistical functional constraint. The statistical functional constraint shall be in terms of cumulative hazard functions as well. Similar to the statistical functional in terms of distribution functions, we can define the statistical functional in terms of cumulative hazard functions as follows.

Definition Suppose $(T_1,\delta_1),\ldots,(T_n,\delta_n)$ are n i.i.d. random vectors as defined in (2.30). Let $U_n = U_n((T_1,\delta_1),\ldots,(T_n,\delta_n))$ be a statistic. If U_n can be written as a functional T of the Nelson-Aalen estimator, $U_n = T(\hat{\Lambda}_{NA})$, where T does not depend on n, then T is called a statistical functional in terms of cumulative hazard functions.

The domain of definition of T is assumed to contain the Nelson-Aalen estimator for all $n \geq 1$, as well as the population cumulative hazard function Λ. Unless otherwise specified, the range of T will be the set of real numbers. The parameter to be estimated is $T(\Lambda)$.

For example, let $X_1, \ldots, X_n$ be n i.i.d. random variables with survival probability $S(t) = P(X > t)$ denoting lifetimes and $C_1, \ldots, C_n$ be n i.i.d. random variables with survival probability $G(t) = P(C > t)$ denoting censoring times. X and C are independent. And only censored observations are available to us.

$$T_i = \min(X_i, C_i), \quad \delta_i = I(X_i \leq C_i), \quad i = 1, \ldots, n \tag{2.60}$$

Suppose there are k distinct uncensored lifetimes $t_1 < t_2 < \ldots < t_k$. Corresponding to each t_i we have n_i, the number of individuals at risk prior to time t_i and d_i, the number of deaths at time t_i. Let $g(t)$ be a real valued function of t and $U_n = \sum_{i=1}^{k} g(t_i)\frac{d_i}{n_i}$. For a general cumulative hazard function $\Lambda(t)$, the functional defined by the following

$$T(\Lambda) = \int g(t) d\Lambda(t) \tag{2.61}$$

satisfies $U_n = T(\hat{\Lambda}_{NA})$, since $\hat{\Lambda}_{NA}(t) = \sum_{t_i \leq t} \frac{d_i}{n_i}$. This is called a linear statistical functional in terms of the cumulative hazard functions.
In particular, suppose we are interested in getting a 95% confidence interval for the cumulative hazard at time t_0, $\Lambda_0(t_0)$, where Λ_0 is the true cumulative hazard function. Hence $\theta_0 = \Lambda_0(t_0)$. In this case, the function g is an indicator function: $g(t) = I[t \leq t_0]$.

Any statistical functional in terms of the cumulative hazard functions but not of the form $\int g(t) d\Lambda(t)$ is called a non-linear statistical functional in terms of the cumulative hazard functions.

For example, let $g(x, y)$ be a function of both x and y, and $U_n = \sum_{i=1}^{k} g(t_i, \sum_{t_j \le t_i} \frac{d_j}{n_j}) \frac{d_i}{n_i}$, for any general cumulative hazard function Λ, the statistical functional T defined as follows

$$T(\Lambda) = \int g\left(t, \Lambda(t)\right) d\Lambda(t) \tag{2.62}$$

satisfies $U_n = T(\hat{\Lambda}_{NA})$, which is called a non-linear statistical functional in terms of the cumulative hazard functions.

In particular, suppose we are interested in getting a 95% confidence interval for the mean of a continuous distribution with cumulative hazard function $\Lambda_0(t)$. Hence $\theta_0 = \int te^{-\Lambda_0(t)} d\Lambda_0(t)$. This formula can be easily verified by the relationship between distribution functions and hazard functions. The mean of a continuous distribution function F_0 is $\int tdF_0(t)$. The relationship between a continuous distribution function F_0 and a continuous cumulative hazard function is $1 - F_0(t) = e^{-\Lambda_0(t)}$ and $d\Lambda_0(t) = \frac{dF_0(t)}{1-F_0(t)}$. Therefore $\int tdF_0(t)$ can be rewritten as $\int te^{-\Lambda_0(t)} d\Lambda_0(t)$. In this case, the function g is $g(t, \Lambda(t)) = te^{-\Lambda(t)}$. It can be shown that $g(t, \Lambda(t)) = te^{-\Lambda(t)}$ satisfies three assumptions of Theorem (3.2.2) in Chapter 3, which implies that g is Hadamard differentiable at Λ_0. We will introduce the Hadamard differentiability later in this section.

For the higher dimension of domain e.g. $\mathbb{D}[0, \tau] \times \mathbb{D}[0, \tau]$, where $\mathbb{D}[0, \tau]$ is the set of the real valued cadlag functions on $[0, \tau]$. The statistical functional defined on $\mathbb{D}[0, \tau] \times \mathbb{D}[0, \tau]$ as

$$T(\Lambda_1(t), \Lambda_2(s)) = \int g_1(t) d\Lambda_1(t) + \int g_2(s) d\Lambda_2(s) \tag{2.63}$$

is called a linear statistical functional in terms of the cumulative hazard functions.

Any statistical functional not of the above form is called a non-linear statistical functional in terms of the cumulative hazard functions, such as

$$T(\Lambda_1(t), \Lambda_2(s)) = \int\int H(t,s)d\Lambda_1(t)d\Lambda_2(s) \tag{2.64}$$

or

$$T(\Lambda_1(t), \Lambda_2(s)) = \int\int H(t,s,\Lambda_1(t),\Lambda_2(s))d\Lambda_1(t)d\Lambda_2(s) \tag{2.65}$$

$\blacksquare$

Next, we briefly introduce three distinct differentiability of statistical functionals in terms of cumulative hazard functions.

Frechet Derivative

The common definition of Frechet differentiability in a normed vector space is as follows.

Definition Let T be a functional

$$T : \mathbb{D} \longrightarrow \mathbb{R} \tag{2.66}$$

where $\mathbb{D}$ is a normed linear space equipped with norm $\|\cdot\|$ and $\mathbb{R}$ is the real line. T is Frechet differentiable at $F \in \mathbb{D}$ if there exists a linear functional $T_F^{'} : \mathbb{D} \longrightarrow \mathbb{R}$ such that,

$$\lim_{\|G-F\|\to 0} \frac{\left|T(G) - T(F) - T_F^{'}(G-F)\right|}{\|G-F\|} = 0, \quad for\ G \in \mathbb{D} \tag{2.67}$$

The linear functional $T_F^{'}$ is called the Frechet derivative of T at F.

Remark In particular, F and G may not be distribution functions but cumulative hazard functions.

Hadamard Derivative

The Frechet differentiability is often too strong and a lot of statistics are not Frechet differentiable (See Fernholz(1983) Example2.3.2). Therefore a weaker form of derivative, Hadamard derivative, is often used, which is defined as follows.

Definition Let T be a functional

$$T : \mathbb{D} \longrightarrow \mathbb{R} \tag{2.68}$$

where $\mathbb{D}$ is a normed linear space equipped with norm $\|\cdot\|$ and $\mathbb{R}$ is the real line. T is Hadamard differentiable at $\theta \in \mathbb{D}$; if $\exists$ a linear functional $T_\theta^{'} : \mathbb{D} \longrightarrow \mathbb{R}$, for any $\delta_n \to 0 \;\; as \;\; n \to \infty$, $D, D_1, D_2, \ldots \in \mathbb{D}$, s.t. $\|D_n - D\| \to 0$, we have

$$\lim_{n\to\infty} \left(\frac{T(\theta + \delta_n D_n) - T(\theta)}{\delta_n} - T_\theta^{'}(D) \right) = 0 \tag{2.69}$$

and $T_\theta^{'}$ is called Hadamard derivative of T at θ.

Remark In particular, θ can be a cumulative hazard function and $\mathbb{D}$ can be $\mathbb{D}[0, \tau]$, which is the set of all real-valued cadlag functions on $[0, \tau]$, where τ is a fixed number. We will show an example of non-linear but Hadamard differentiable statistical functional in terms of cumulative hazard functions in Theorem 3.2.2 in Chapter 3.

Gateaux Derivative

An even weaker derivative is called Gateaux derivative which is defined as below.

Definition Let T be a functional

$$T : \mathbb{D} \longrightarrow \mathbb{R} \tag{2.70}$$

where $\mathbb{D}$ is normed linear space equipped with norm $\|\cdot\|$ and $\mathbb{R}$ is the real line. T is Gateaux differentiable at F in the direction G if the following limit exists

$$T_F'(G) = \lim_{\epsilon \to 0} \frac{T((1-\epsilon)F + \epsilon G) - T(F)}{\epsilon} \tag{2.71}$$

and if the limits exist for all $G \in \mathbb{D}$, then we say that T is Gateaux differentiable at F.

Remark In particular, F and G may not be distribution functions but cumulative hazard functions.

A uniform definition of three distinct differentiability can be found in Fernholz (1983) Chapter 3.

Definition (Fernholz (1983)) Let $\mathbb{V}$ and $\mathbb{W}$ be topological vector spaces and let $L(\mathbb{V}, \mathbb{W})$ be the set of continuous linear transformations from $\mathbb{V}$ to $\mathbb{W}$. Let $\mathbb{S}$ be a class of subsets of $\mathbb{V}$ such that every subset consisting of a single point belongs to $\mathbb{S}$, and let $\mathbb{A}$ be an open subset of $\mathbb{V}$.

A function

$$T: \quad \mathbb{A} \longrightarrow \mathbb{W} \tag{2.72}$$

is $\mathbb{S}$-differentiable at $F \in \mathbb{A}$, if there exists $T_F' \in L(\mathbb{V}, \mathbb{W})$ such that for any $K \in \mathbb{S}$

$$\lim_{t \to 0} \frac{T(F + tH) - T(F) - T_F'(tH)}{t} = 0 \tag{2.73}$$

uniformly for $H \in K$. The linear function T_F' is called the $S-derivative$ of T at F. It is convenient to define the remainder term

$$R(T, F, H) = T(F + H) - T(F) - T_F'(H) \tag{2.74}$$

With this notation (2.73) is equivalent to : for any neighborhood N of 0 in $\mathbb{W}$. there exists $\epsilon > 0$ such that if $|t| < \epsilon$ then

$$\frac{R(T, F, tH)}{t} \in N \tag{2.75}$$

for all $H \in K$.

Here are three particular types of differentiation that we interested in:

- $\mathbb{S} = \{bounded\ subsets\ of\ \mathbb{V}\}$, this corresponds to Frechet differentiation.
- $\mathbb{S} = \{compact\ subsets\ of\ \mathbb{V}\}$, this corresponds to Hadamard differentiation.
- $\mathbb{S} = \{single\ point\ subsets\ of\ \mathbb{V}\}$, this corresponds to Gateaux differentiation.

It is clear from the above uniform definition that Frechet differentiability implies Hadamard differentiability and Hadamard differentiability implies Gateaux differentiability. ∎

In summary, it is much more convenient to analyze the right censored data by using hazard functions than by using distribution functions. First, an explicit form of maximum can be obtained when we use a constraint in terms of cumulative hazard functions while no explicit form of maximum can be obtained using a constraint in terms of distribution functions. Second, the Nelson-Aalen estimator, which is the nonparametric maximum likelihood estimator of the cumulative hazard functions, can be expressed in a form of martingale and the predictable integration with respect to the Nelson-Aalen estimator is also a martingale. With the expression of martingales, the Martingale Central Limit Theorem can be applied to obtain their asymptotic properties. However, the Kaplan-Meier estimator, which can be considered as the NPMLE of distribution functions, can not be expressed as a simple form of martingale as the Nelson-Aalen estimator does. Without the convenient martingale expression, it is difficult to analyze the right censored data by using distribution functions.

In result, to analyze the empirical likelihood with the right censored data subject to a non-linear statistical functional, we shall consider the statistical functional in terms of cumulative hazard functions. In the following chapters, we shall investigate the Hadamard differentiability of the non-linear statistical functional in terms of cumulative hazard functions.

Chapter 3 Empirical Likelihood Ratio Subject to Nonlinear Statistical Functional in Terms of Cumulative Hazard with Right Censored Data

3.1 Introduction

Background

In this chapter, we prove that $-2\log ELR(\theta_0)$ converges to a $\chi^2_{(1)}$, when the following null hypothesis is true,

$$H_0 \; : \; \int g\left(t, \Lambda(t)\right) d\Lambda(t) = \theta_0 \tag{3.1}$$

$T(\Lambda) = \int g(t, \Lambda(t))d\Lambda(t)$ is a non-linear but Hadamard differentiable statistical functional. The sufficient conditions for Hadamard differentiability are listed in Theorem 3.2.2. $ELR(\theta_0)$ is the maximum of the empirical likelihood ratio (ELR) function subject to the non-linear statistical functional constraint $\int g(t, \Lambda(t))d\Lambda(t) = \theta_0$, where $\theta_0 = \int g(t, \Lambda_0(t))d\Lambda_0(t)$ and Λ_0 is the true cumulative hazard function. The ELR function is defined as follows.

$$ELR = \frac{EL(\Lambda)}{EL(\hat{\Lambda}_{NA})} \tag{3.2}$$

where $EL(\cdot)$ is defined in (2.15) and $\hat{\Lambda}_{NA}$ is the Nelson-Aalen estimator.

The empirical likelihood method was first proposed by Thomas and Grunkemeier (1975). They heuristically prove that the empirical likelihood ratio statistic for a survival probability has a χ^2_1 limiting distribution under the null hypothesis that $P(X > a) = p_0$, where a is a fixed real number and p_0 is a hypothesized probability. Owen (1988,1990) and many others have developed the empirical likelihood into a general methodology but for uncensored data. Owen (1988) proves that $-2\log ELR(\theta_0)$ converges to a χ_1^2 subject to a linear statistical functional constraint in terms of distribution functions. A direct generalization of Owen's setting to the right censored data is difficult, since there is no explicit maximization form of Lagrange multiplier method. Pan and Zhou (2002) generalize Owen's setting to the right censored data using a linear statistical functional constraint in terms of cumulative hazard. In this chapter, we generalize Pan and Zhou's results to a nonlinear statistical functional constraint in terms of cumulative hazard functions as follows.

$$\int g(t, \Lambda(t))d\Lambda(t) = \theta_0 \tag{3.3}$$

We believe that analyzing this kind of hazard-type constraint is a valuable theoretical contribution in its own right. A lot of constraints in terms of distribution functions with right censored data that are difficult to analyze now can be solved by transforming the constraint to the form (3.3). Furthermore, the method introduced in this chapter can be easily applied to two sample problems. We will discuss two sample problems in chapter 5.

Motivation

Our motivation to analyze this kind of constraint is if we can deal with the constraint of the following form

$$\int g(t, \Lambda(t))d\Lambda(t) = \theta_0 \tag{3.4}$$

then we can deal with any constraint that can be transformed to form (3.4) which are intricate originally, such as the hypothesis testing of the mean $\int g(t)dF(t) = \mu$ and the hypothesis testing of the Gini index.

To be more specific, suppose the hypothesis testing of the mean is

$$\int g(t)dF(t) = \mu_0 \tag{3.5}$$

Assuming the true distribution is continuous, by the relationship between distribution functions and hazard functions $d\Lambda(t) = \frac{dF(t)}{1-F(t)}$, $1 - F(t) = \exp(-\Lambda(t))$, the constraint in terms of hazard is

$$\int g(t)e^{-\Lambda(t)}d\Lambda(t) = \mu_0 \tag{3.6}$$

which is of the form of our new generalization.

Next, we briefly introduce the Gini index and see how we can do hypothesis testing of Gini index using a non-linear statistical functional constraint in terms of hazard functions.

Corrado Gini presents the index, which is known as "Gini index" today, for the first time in 1912 in his book "Variability and Mutability". The Gini index can be used to measure the dispersion of a distribution of income, or consumption, or wealth with the most widely use on the dispersion of income. Therefore, in this dissertation, we shall focus on the Gini index in the context of income distribution. The usual definition of Gini index is as follows.

Definition The Gini index is a measure of statistical dispersion intended to represent the income distribution of a nation's residents. This is the most commonly used measure of inequality. The coefficient varies between 0, which reflects complete equality and 1, which indicates complete inequality (one person has all the income while all others have none). ∎

Generally, there are two different approaches to analyze the Gini index. One is based on discrete distributions and the other is based on continuous distributions. The difference between these two approaches is that the discrete approach assumes that the population is finite while the continuous approach assumes that the population is infinite. In this dissertation, we shall focus on the continuous approach.

The Gini index has many interesting formulations and interpretations. We briefly introduce two different approaches: **Geometric Approach** and **Gini's Mean Difference Approach** in the following. See Xu (2004) and Ceriani and Verme (2001) for more detailed discussion about various formulations and interpretations of Gini index.

Geometric Approach

Figure 3.1 is the graphic representation of Gini index. The x-axis represents the cumulative share of people from lowest to highest income and the y-axis is the cumulative share of income earned. The line at 45 degree represents perfect equality

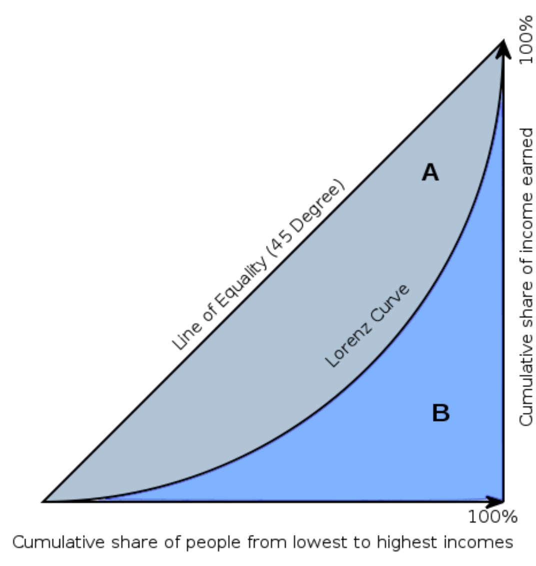

Figure 3.1: Graphic Representation of Gini Coefficient

of income, since $x\%$ of the total income of the population is cumulatively earned by the bottom $x\%$ of the population. But in reality, this is not the case. In reality, this relationship is depicted by a curve called **Lorenz Curve**. In reality, for example, bottom 50% of the population might just earn 20% of the total income. The Gini index can then be considered as the ratio of the area that lies between the 45 degree line and the Lorenz curve, which is area A, over the total area under the 45 degree line, which is area $A + B$.

$$G = \frac{A}{A + B} \tag{3.7}$$

Gini's Mean Difference Approach

Gini index can also be represented by the so called Gini's (absolute and relative) mean difference. In fact, the Gini index is just the half of the Gini's relative mean difference, which will be explained later.

Gini's absolute mean difference for a continuous income distribution F is defined as follows.

$$D = E\left|X - Y\right| = \int_0^{+\infty}\int_0^{+\infty} |x - y| dF(x) dF(y) \tag{3.8}$$

The value of D is the average absolute difference of income of two randomly selected individuals and reflects the income inequality in the population. It is straightforward to see that $0 \leq D \leq 2\mu$, where $\mu = E(Y) = \int_0^{+\infty} y dF(y)$ is the population mean income. The Gini index is defined as the normalized mean difference

$$G = \frac{D}{2\mu} \tag{3.9}$$

so $G \in [0, 1]$.

Since we have

$$\begin{aligned}
D &= \int_0^{+\infty}\int_0^{+\infty} |x-y|\, dF(x)dF(y) \\
&= \int_0^{+\infty}\int_0^{y} (y-x)dF(x)dF(y) + \int_0^{+\infty}\int_y^{+\infty} (x-y)dF(x)dF(y) \\
&= \int_0^{+\infty} yF(y)dF(y) - \int_0^{+\infty}\int_0^{y} xdF(x)dF(y) + \int_0^{+\infty}\int_y^{+\infty} xdF(x)dF(y) \\
&- \int_0^{+\infty} y(1-F(y))dF(y) \\
&= \int_0^{+\infty} y(2F(y)-1)dF(y) + \int_0^{+\infty}\int_y^{+\infty} xdF(x)dF(y) - \int_0^{+\infty}\int_0^{y} xdF(x)dF(y) \\
&= \int_0^{+\infty} y(2F(y)-1)dF(y) + \int_0^{+\infty} x\int_0^{x} dF(y)dF(x) - \int_0^{+\infty} x\int_x^{+\infty} dF(y)dF(x) \\
&= \int_0^{+\infty} y(2F(y)-1)dF(y) + \int_0^{+\infty} xF(x)dF(x) - \int_0^{+\infty} x(1-F(x))dF(x) \\
&= \int_0^{+\infty} y(2F(y)-1)dF(y) + \int_0^{+\infty} x(2F(x)-1)dF(x) \\
&= 2\int_0^{+\infty} y(2F(y)-1)dF(y)
\end{aligned} \tag{3.10}$$

the Gini index can be represented by the statistical functional of distribution function F as follows, where F is the income cumulative distribution function of a nation's population.

$$G = \frac{\int_0^{+\infty} y(2F(y)-1)dF(y)}{\mu} \tag{3.11}$$

where $\mu = \int_0^{\infty} ydF(y)$ is the expected total income.

Suppose we would like to do the hypothesis testing of Gini index. The null hypothesis is

$$H_0 : G = \mu_0$$

Similar to the hypothesis testing of the mean, we assume the true distribution is continuous, then the null hypothesis H_0 can be transformed to

$$\int_0^{+\infty} \left((1-\mu_0)te^{-\Lambda(t)} - 2te^{-2\Lambda(t)}\right) d\Lambda(t) = 0 \tag{3.12}$$

which is a nonlinear statistical functional of the cumulative hazard function Λ.

This test of $H_0 : G = \mu_0$ can also be formulated in terms of cumulative distribution functions. This also serves as a motivation example for the Chapter 4.

A simulation of Gini's absolute mean difference (D) and a discussion of the variance estimation of the Gini index will be presented in **Simulation 6** and **Simulation 7** of Chapter 6 respectively. In Simulation 7, we will compare the coverage probability and average length of the confidence intervals based on our method and two other empirical likelihood methods. An application of Theorem 4.56 to the Gini's absolute mean difference (D) using real data is presented in **Real Data Analysis** of Chapter 6.

3.2 Lemma and Theorem

First of all, we establish a theorem which is the foundation of all lemmas and theorems later.

Theorem 3.2.1 *Suppose we have two statistical functional constraints in terms of cumulative hazard functions as follows*

$$T_1(\Lambda) = \theta_0, \quad T_2(\Lambda) = \theta_0 \tag{3.13}$$

which satisfy $\theta_0 = T_1(\Lambda_0)$ *and* $\theta_0 = T_2(\Lambda_0)$*.*

If for any statistic $\hat{\Lambda}_n(t)$*, s.t.* $\left\|\hat{\Lambda}_n(t) - \Lambda_0(t)\right\| = O_p\left(\frac{1}{\sqrt{n}}\right)$*, we have*

$$\left|T_1(\hat{\Lambda}_n) - T_2(\hat{\Lambda}_n)\right| = o_p\left(\frac{1}{\sqrt{n}}\right) \tag{3.14}$$

then we have

$$-2\log ELR_1(\theta_0) + 2\log ELR_2(\theta_0) \xrightarrow{P} 0, \ \textit{as } n \to \infty \tag{3.15}$$

where $ELR_1(\theta_0)$ *and* $ELR_2(\theta_0)$ *are the maximums of the ELR function (3.2) subject to the corresponding constraint* $T_1(\Lambda) = \theta_0$ *and* $T_2(\Lambda) = \theta_0$ *respectively.*

Proof See Pan and Zhou (2002) for the proof of the theorem. ∎

There is no explicit maximum of ELR function (3.2) subject to a non-linear statistical functional constraint. However, the Hadamard derivative of the non-linear statistical functional is linear. We will prove this later. Moreover, the difference between the non-linear statistical functional and its Hadamard derivative is $o_p\left(\frac{1}{\sqrt{n}}\right)$ in the domain of some statistics. By Theorem (3.2.1), as long as we prove that

$$-2\log ELR(\theta_0) \xrightarrow{D} \chi^2_{(1)} \tag{3.16}$$

under the null hypothesis of the linear statistical functional, which is the Hadamard derivative of the non-linear statistical functional, so do we have (3.16) under the null hypothesis of the non-linear statistical functional. Therefore, we investigate the Hadamard differentiability of the non-linear statistical functional in the following theorem.

In the following theorem, we prove that, under some regularity conditions, the non-linear statistical functional $T(\Lambda) = \int g(t, \Lambda(t))d\Lambda(t)$ is Hadamard differentiable and the Hadamard derivative of T at Λ_0 is a linear statistical functional. And the difference between the non-linear and linear statistical functional is $o_p(\frac{1}{\sqrt{n}})$ in the domain of some statistics.

Theorem 3.2.2 *Let $T(\Lambda) = \int g(t, \Lambda(t))d\Lambda(t)$ be a non-linear statistical functional defined on $\mathbb{D}[0, \tau]$. $\mathbb{D}[0, \tau]$ is the set of all real valued cadlag functions on $[0, \tau]$ and equipped with the sup norm:*

$$\|f\| = \sup_{x \in [0,\tau]} |f(x)| \tag{3.17}$$

Define $h = \frac{\partial g}{\partial \Lambda}$ and $\tilde{h}(t) = \int_t^{+\infty} h(s, \Lambda_0(s))d\Lambda_0(s)$, where $\Lambda_0(t)$ is a continuous cumulative hazard function. Under some regularity conditions,

Assumption (A) *$g(t, \Lambda(t))$ is left continuous with respect to t and twice differentiable with respect to $\Lambda(t)$.*

Assumption (B) *$|g\left(t, \Lambda(t)\right)| \leq A(t),\ |h\left(t, \Lambda(t)\right)| \leq B(t)$ for all $t \in [0, \tau]$ and $\Lambda \in \mathbb{D}[0, \tau]$, where $A(t)$ is integrable with respect to any cadlag function $D(t) \in \mathbb{D}[0, \tau]$ and $B(t)$ is integrable with respect to $\Lambda_0(t)$.*

Assumption (C) *$\Lambda(0) = 0$ and $\Lambda(\tau) \leq M$ for some $M \in \mathbb{R}$.*

T is Hadamard differentiable at Λ_0 with derivative

$$T'_{\Lambda_0}\left(\Lambda(t) - \Lambda_0(t)\right) = \int \left(g(t, \Lambda_0(t)) + \tilde{h}(t)\right) d\left(\Lambda(t) - \Lambda_0(t)\right) \tag{3.18}$$

And the remaining term is:

$$\left|T(\Lambda(t)) - T(\Lambda_0(t)) - T'_{\Lambda_0}(\Lambda(t) - \Lambda_0(t))\right| = o\left(\|\Lambda(t) - \Lambda_0(t)\|\right) \tag{3.19}$$

In particular, if $\Lambda(t) = \hat{\Lambda}_{NA}(t)$*, the remaining term is* $o_p\left(\frac{1}{\sqrt{n}}\right)$*. We will show in Lemma 3.2.3 that if* $\Lambda(t)$ *deviates a little from* $\hat{\Lambda}_{NA}(t)$*, the remaining term is still* $o_p(\frac{1}{\sqrt{n}})$*.*

Proof We begin the proof with a discussion of functions g, h and $\tilde{h}$.

First We shall point out that the function g must be a function of both t and Λ. If g only depends on t, it is just Pan and Zhou (2002)'s setting. If g only depends on Λ, the integration $\int g(\Lambda)d\Lambda$ will be a fixed number thus we cannot put a constraint on it. In this case, we cannot do hypothesis testing or generate confidence intervals.

Second h is the partial derivative of g with respect to Λ. For example, if $g(t, \Lambda(t)) = te^{-\Lambda(t)}$, then $h(t, \Lambda(t)) = -te^{-\Lambda(t)}$.

Third Efron and Johnstone (1990) define the advanced-time transformation $\tilde{g}(t)$ for a function $g(t)$ with respect to a continuous cumulative distribution function $F_0(t)$ as

$$\tilde{g}(t) = \frac{\int_t^\infty g(s)dF_0(s)}{1 - F_0(t)} = E_{F_0}[g(X)|X > t] \tag{3.20}$$

From (3.20), it is clear that the advanced-time transformation $\tilde{g}(t)$ is the conditional expectation of $g(X)$ given $X > t$.

Parallelly, we can define the advanced-time transformation $\tilde{h}(t)$ for a function h(t) with respect to a continuous cumulative hazard function $\Lambda_0(t)$ as

$$\tilde{h}(t) = \int_t^\infty h(s)d\Lambda_0(s) \tag{3.21}$$

Now for every $\tilde{g}$, there exists a $\tilde{h}$, such that $\tilde{g}(t)\,(1 - F_0(t)) = \tilde{h}(t)$. This fact is easy to see if we choose $h(t) = g(t)e^{-\Lambda_0(t)}$.

Fourth Since g is left continuous with respect to t and twice differentiable with respect to Λ (Assumption (A)), it is obvious that $h = \frac{\partial g}{\partial \Lambda}$ is also left continuous with respect to t and differentiable with respect to Λ. Therefore $\tilde{h}(t) = \int_t^{\infty} h(s, \Lambda_0(s))d\Lambda_0(s)$ is left continuous with respect to t, which means $\tilde{h}$ is predictable.

For example, let $g(t, \Lambda(t)) = \left((1-\mu_0)te^{-\Lambda(t)} - 2te^{-2\Lambda(t)}\right)$, which is from the hypothesis testing of Gini coefficient, $G = \mu_0$ $(0 \le \mu_0 \le 1)$. Obviously, g is left continuous with respect to t and twice differentiable with respect to Λ. g is also bounded by $A(t) = 3t$, which is integrable with respect to any cadlag function in $\mathbb{D}[0, \tau]$. $h(t, \Lambda(t)) = \frac{\partial g}{\partial \Lambda} = -(1-\mu_0)te^{-\Lambda(t)} + 4te^{-2\Lambda(t)}$ is bounded by $B(t) = 5t$, which is integrable with respect to Λ_0 in $[0, \tau]$. Suppose the true distribution is exp(1) and we would like to test the null hypothesis, $G = 0.5$. The advanced-time transformation $\tilde{h}(t)$ for $h(t, \Lambda_0(t))$ with respect to the continuous cumulative hazard function $\Lambda_0(t) = t$ is as follows.

$$\begin{aligned} \tilde{h}(t) &= \int_t^{\infty} h(s, \Lambda_0(s))d\Lambda_0(s) = \int_t^{\infty} \left(-0.5se^{-s} + 4se^{-2s}\right) ds \\ &= -\frac{1}{2}\left(t+1\right)e^{-t} + \left(2t+1\right)e^{-2t} \end{aligned} \quad (3.22)$$

∎

Next, we study the Hadamard differentiability of the non-linear statistical functional $T(\Lambda) = \int g(t, \Lambda(t))d\Lambda(t)$.

By the definition of Hadamard differentiability, we need to prove that for any $D_1, D_2, \ldots,$ and $D \in \mathbb{D}$, such that $\|D_n - D\| \to 0$ as $n \to \infty$ and $\delta_n \to 0$ as $n \to \infty$, we have

$$\lim_{n\to\infty} \left(\frac{T(\Lambda_0 + \delta_n D_n) - T(\Lambda_0)}{\delta_n} - T'_{\Lambda_0}(D) \right) = 0 \quad (3.23)$$

According to the definition of linear statistical functional in terms of the cumulative hazard functions, we notice that T'_{Λ_0} is a linear statistical functional. Therefore, we have $\delta_n T'_{\Lambda_0}(D) = T'_{\Lambda_0}(\delta_n D)$. Then we have the following.

$$
\begin{aligned}
&\lim_{n\to\infty}\left(\frac{T(\Lambda_0+\delta_n D_n)-T(\Lambda_0)}{\delta_n}-T'_{\Lambda_0}(D)\right)\\
&=\lim_{n\to\infty}\frac{T(\Lambda_0+\delta_n D_n)-T(\Lambda_0)-T'_{\Lambda_0}(\delta_n D)}{\delta_n}\\
&=\lim_{n\to\infty}\frac{\int g(t,\Lambda_0+\delta_n D_n)d(\Lambda_0+\delta_n D_n)-\int g(t,\Lambda_0)d\Lambda_0-\int(g(t,\Lambda_0)+\tilde{h}(t))d(\delta_n D)}{\delta_n}\\
&=\lim_{n\to\infty}\int\frac{g(t,\Lambda_0+\delta_n D_n)-g(t,\Lambda_0)}{\delta_n}d\Lambda_0\\
&+\lim_{n\to\infty}\int\left(g(t,\Lambda_0+\delta_n D_n)-g(t,\Lambda_0)\right)d(D_n-D)\\
&+\lim_{n\to\infty}\int g(t,\Lambda_0)d(D_n-D)+\lim_{n\to\infty}\int\left(g(t,\Lambda_0+\delta_n D_n)-g(t,\Lambda_0)\right)dD-\int\tilde{h}(t)dD\\
&=(1)+(2)+(3)+(4)-(5)
\end{aligned}
\tag{3.24}
$$

We prove that (1)=(5) and (2)=(3)=(4)=0 in the following. To prove these, **Dominated Convergence Theorem** will be applied. See Appendix for the **Dominated Convergence Theorem**. To apply the Dominated Convergence Theorem, we need the **Assumption (B)**. By the Dominated Convergence Theorem, we can switch the limitation and the integration.

By the L'Hospital's Rule, we have

$$
\begin{aligned}
&\lim_{n\to\infty}\frac{g\left(t,\Lambda_0+\delta_n D_n\right)-g\left(t,\Lambda_0\right)}{\delta_n}=\lim_{n\to\infty}\frac{\frac{d(g(t,\Lambda_0+\delta_n D_n)-g(t,\Lambda_0))}{d(\delta_n)}}{\frac{d(\delta_n)}{d(\delta_n)}}\\
&=\lim_{n\to\infty}h\left(t,\Lambda_0+\delta_n D_n\right)D_n(t)=h\left(t,\Lambda_0\right)D(t)
\end{aligned}
\tag{3.25}
$$

$g\left(t,\Lambda_0+\delta_n D_n\right)$ can be considered as a function of δ_n, the derivative of $g\left(t,\Lambda_0+\delta_n D_n\right)$ with respect to δ_n is $h\left(t,\Lambda_0+\delta_n D_n\right)D_n$. By Fernholz (1983) Lemma 4.4.1, we learn that $D_n(t)$ is uniformly bounded, which means $D_n(t)$ is bounded by a function that does not depend on n. Since $|h(t, \Lambda(t))| \leq B(t)$, $\forall\ t \in [0,\tau]$ and $\forall\ \Lambda(t) \in \mathbb{D}[0,\tau]$, $h(t, \Lambda_0 + \delta_n D_n)\, D_n(t)$ is uniformly bounded as well. Because $g(t, \Lambda)$ is twice differentiable with respect to Λ, it is easy to show that $g(t, \Lambda_0 + xD_n)$ is continuous with respect to x on $[0, \delta_n]$ and differentiable with respect to x on $(0, \delta_n)$. By **Mean Value Theorem**, there exists a $\xi_n \in [0, \delta_n]$, s.t.

$$\frac{g(t, \Lambda_0 + \delta_n D_n) - g(t, \Lambda_0)}{\delta_n} = h(t, \Lambda_0 + \xi_n D_n)\, D_n \tag{3.26}$$

$h(t, \Lambda_0 + \xi_n D_n)\, D_n$ is uniformly bounded by some measureable integrable function, consequently, $\frac{g(t,\Lambda_0+\delta_n D_n)-g(t,\Lambda_0)}{\delta_n}$ is uniformly bounded as well.
By the Dominated Convergence Theorem, we have

$$\begin{aligned}(1) &= \lim_{n\to\infty} \int \frac{g(t, \Lambda_0 + \delta_n D_n) - g(t, \Lambda_0)}{\delta_n} d\Lambda_0 = \int \lim_{n\to\infty} \frac{g(t, \Lambda_0 + \delta_n D_n) - g(t, \Lambda_0)}{\delta_n} d\Lambda_0 \\ &= \int h(t, \Lambda_0(t)) D(t) d\Lambda_0(t)\end{aligned} \tag{3.27}$$

Next we prove that (5) = (1).

$$\begin{aligned}(5) &= \int \tilde{h}(t) dD(t) = \int \int_t^\infty h(s, \Lambda_0(s)) d\Lambda_0(s) dD(t) \\ &= \int \left(\int_0^s dD(t) \right) h(s, \Lambda_0(s)) d\Lambda_0(s) = \int h(s, \Lambda_0(s)) D(s) d\Lambda_0(s) = (1)\end{aligned} \tag{3.28}$$

Since

$$|D_n(t) - D(t)| \leq \sup_{t\in[0,\tau]} |D_n(t) - D(t)| = \|D_n - D\| \to 0,\ as\ \ n \to \infty, \forall t \in [0,\tau] \tag{3.29}$$

In particular, we have

$$\begin{aligned}\lim_{n\to\infty}|D_n(\tau)-D(\tau)|=0\\ \lim_{n\to\infty}|D_n(0)-D(0)|=0\end{aligned} \tag{3.30}$$

By the integration by parts, we have

$$\begin{aligned}(3) &= \lim_{n\to\infty}\int_0^\tau g(t,\Lambda_0(t))d(D_n(t)-D(t))\\ &= \lim_{n\to\infty}\left(g(t,\Lambda_0(t))(D_n(t)-D(t))|_0^\tau - \int (D_n(t)-D(t))dg\left(t,\Lambda_0(t)\right)\right)\\ &= \lim_{n\to\infty} g(t,\Lambda_0(t))(D_n(t)-D(t))|_0^\tau - \lim_{n\to\infty}\int (D_n(t)-D(t))dg\left(t,\Lambda_0(t)\right)\end{aligned} \tag{3.31}$$

We have

$$\left|\int \left(D_n(t)-D(t)\right)dg(t,\Lambda_0(t))\right| \le \int \left|D_n(t)-D(t)\right| d\left|g(t,\Lambda_0(t)\right| \tag{3.32}$$

D_n and D are uniformly bounded, so is $D_n - D$.

By the Dominated Convergence Theorem

$$\lim_{n\to\infty}\int \left|D_n(t)-D(t)\right| d\left|g(t,\Lambda_0(t))\right| = \int \lim_{n\to\infty}\left|D_n(t)-D(t)\right| d\left|g(t,\Lambda_0(t))\right| = 0 \tag{3.33}$$

Therefore

$$\lim_{n\to\infty}\left|\int \left(D_n(t)-D(t)\right)dg(t,\Lambda_0(t))\right| \le \lim_{n\to\infty}\int \left|D_n(t)-D(t)\right| d\left|g(t,\Lambda_0(t))\right| = 0 \tag{3.34}$$

Now we prove that $(3) = 0$.

Since $|g(t, \Lambda(t))| \leq A(t)$ *holds for any t and* $\Lambda(t)$, we have

$$\begin{aligned}
&\left|\int (g(t,\Lambda_0+\delta_n D_n)-g(t,\Lambda_0))d\left(D_n-D\right)\right| \\
&\leq \int \left|g(t,\Lambda_0+\delta_n D_n)-g(t,\Lambda_0)\right| d\left|D_n-D\right| \\
&\leq \int \left(\left|g(t,\Lambda_0+\delta_n D_n)\right|+\left|g(t,\Lambda_0)\right|\right) d\left|D_n-D\right| \leq 2\int A(t) d\left|D_n(t)-D(t)\right|
\end{aligned} \tag{3.35}$$

By the integration by parts and Dominated Convergence Theorem, we have

$$\begin{aligned}
&\lim_{n\to\infty}\int_0^\tau A(t)d\left|D_n(t)-D(t)\right| \\
&=\lim_{n\to\infty}\left(A(t)\left|D_n(t)-D(t)\right|\big|_0^\tau-\int \left|D_n(t)-D(t)\right| dA(t)\right) \\
&=\lim_{n\to\infty}A(t)\left|D_n(t)-D(t)\right|\big|_0^\tau-\int \lim_{n\to\infty}\left|D_n(t)-D(t)\right| dA(t)=0
\end{aligned} \tag{3.36}$$

Now we have the following

$$\begin{aligned}
&|(2)|=\lim_{n\to\infty}\left|\int (g(t,\Lambda_0+\delta_n D_n)-g(t,\Lambda_0))d(D_n-D)\right| \\
&\leq 2\lim_{n\to\infty}\int A(t)d\left|D_n(t)-D(t)\right|=0
\end{aligned} \tag{3.37}$$

so we prove that (2)=0.

Again, since $|g(t,\Lambda_0+\delta_n D_n)-g(t,\Lambda_0)| \leq 2A(t)$, by the Dominated Convergence Theorem, we have,

$$\begin{aligned}
|(4)|&=\lim_{n\to\infty}\left|\int \left(g(t,\Lambda_0+\delta_n D_n)-g(t,\Lambda_0)\right) dD(t)\right| \\
&=\lim_{n\to\infty}\int \left|g(t,\Lambda_0+\delta_n D_n)-g(t,\Lambda_0)\right| d|D(t)| \\
&=\int \lim_{n\to\infty}\left|g(t,\Lambda_0+\delta_n D_n)-g(t,\Lambda_0)\right| d|D(t)|=0
\end{aligned} \tag{3.38}$$

In conclusion, (3.24)=0, which means we prove that

$$\lim_{n\to\infty}\left(\frac{T(\Lambda_0+\delta_n D_n)-T(\Lambda_0)}{\delta_n}-T'_{\Lambda_0}(D)\right)=0 \tag{3.39}$$

In other words, T is Hadamard differentiable at Λ_0 and $T'_{\Lambda_0}(\cdot)$ is the Hadamard derivative of T at Λ_0.

By **Assumption (C)**

$$\Lambda(0)=0 \;\; and \;\; \Lambda(\tau)\le M \;\; for \;\; some \;\; M\in\mathbb{R} \tag{3.40}$$

and Fernholz (1983) **Proposition 4.3.3**, we have

$$\left|T(\Lambda(t))-T(\Lambda_0(t))-T'_{\Lambda_0}(\Lambda(t)-\Lambda_0(t))\right|=o\left(\|\Lambda(t)-\Lambda_0(t)\|\right) \tag{3.41}$$

In particular, if $\Lambda(t)=\hat{\Lambda}_{NA}(t)$, it is well known that $\left\|\hat{\Lambda}_{NA}(t)-\Lambda_0(t)\right\|=O_p(\frac{1}{\sqrt{n}})$. Then the remaining term is $o_p(\frac{1}{\sqrt{n}})$. Now the proof of Theorem (3.2.2) is accomplished. ∎

Before our next lemma, we shall review the empirical likelihood introduced in the Chapter 2.

Suppose that $X_1,\ldots,X_n$ are n i.i.d. nonnegative random variables denoting the lifetimes with a continuous distribution function F_0. Independent of the lifetimes there are n censoring times $C_1,\ldots,C_n$, which are i.i.d. with a distribution function G_0. Only the censored observations are available to us:

$$T_i=\min(X_i,C_i),\;\delta_i=I[X_i\le C_i],\;i=1,2,\ldots,n \tag{3.42}$$

Remember that the empirical likelihood based on censored observations (T_i, δ_i) is

$$EL(F) = \prod_{i=1}^{n} [\Delta F(T_i)]^{\delta_i} [1 - F(T_i)]^{1-\delta_i} \tag{3.43}$$

By the relationship between distribution functions and hazard functions

$$1 - F(t) = \prod_{s \leq t} (1 - \Delta\Lambda(s)) \qquad and \qquad \Delta\Lambda(t) = \frac{\Delta F(t)}{1 - F(t-)} \tag{3.44}$$

(3.43) can be rewritten in terms of cumulative hazard function as follows.

$$EL(\Lambda) = \prod_{i=1}^{n} \left\{ [\Delta\Lambda(T_i)]^{\delta_i} \left[\prod_{j:T_j<T_i} (1 - \Delta\Lambda(T_j)) \right]^{\delta_i} \left[\prod_{j:T_j \leq T_i} (1 - \Delta\Lambda(T_j)) \right]^{1-\delta_i} \right\} \tag{3.45}$$

The hazard function that maximizes the likelihood $EL(\Lambda)$ without any constraint is the Nelson-Aalen estimator (Andersen et. al. (1993)). We denote the Nelson-Aalen estimator as $\hat{\Lambda}_{NA}(t)$.

On the other hand, a simpler version of the likelihood can be obtained if we merge the second and third terms in (3.45) and replace it with $\exp\{-\Lambda(T_i)\}$, which is called a Poisson extension of the likelihood introduced by Murphy (1995):

$$AL(\Lambda) = \prod_{i=1}^{n} [\Delta\Lambda(T_i)]^{\delta_i} \exp\{-\Lambda(T_i)\} \tag{3.46}$$

The above formula of AL is only valid for continuous distributions. In the case of a discrete distribution, the difference is small and negligible when n is large. We will show this fact in Theorem 3.2.6 later in this chapter.

Preceding our next lemma, we point out that the last jump of a proper discrete cumulative hazard function must be one. It is clear from the second equation of (3.44). It is similar to the restriction of the distribution function that all jumps sum to one. And we assume no tie in the uncensored observations. Without loss of generality we assume $T_1 \leq T_2 \leq \ldots \leq T_n$ where ties are only possible between censored observations.

In the next lemma, we will maximize AL (3.46) subject to the linear statistical functional constraint $T(\Lambda_0) + T'_{\Lambda_0}(\Lambda - \Lambda_0) = \theta$. $T'_{\Lambda_0}(\Lambda - \Lambda_0)$ is the Hadamard derivative of $T(\Lambda) = \int g(t, \Lambda(t))d\Lambda(t)$ at Λ_0.

Lemma 3.2.3 *If the constraint below is feasible*

$$T(\Lambda_0) + T'_{\Lambda_0}(\Lambda - \Lambda_0) = \int (g(t, \Lambda_0(t)) + \tilde{h}(t))d\Lambda(t) - \int \tilde{h}(t)d\Lambda_0(t) = \theta \tag{3.47}$$

A discussion of feasibility and the feasible value of θ in the above constraint are given by the interval at the end of the proof.

Then the maximum of AL (3.46) under constraint is obtained when

$$w_i = \Delta\Lambda(T_i) = \Delta\hat{\Lambda}_{NA}(T_i)\frac{1}{1 + \lambda Z_i}, \;\; i = 1, 2, \ldots, n-1 \tag{3.48}$$

where λ is the solution of the following equation.

$$l(\lambda) = \theta \tag{3.49}$$

where

$$\begin{aligned} l(\lambda) &= \sum_{i=1}^{n-1} \Big(g(T_i, \Lambda_0(T_i)) + \tilde{h}(T_i)\Big) \frac{\delta_i}{n-i+1} \frac{1}{1 + \lambda \frac{\delta_i\left(g(T_i,\Lambda_0(T_i)) + \tilde{h}(T_i)\right)}{\frac{n-i+1}{n}}} + g(T_n)\delta_n \\ &= \frac{1}{n}\sum_{i=1}^{n-1} \frac{\delta_i Z_i}{1 + \lambda Z_i} + \frac{1}{n}\delta_n Z_n - \hat{\theta} \end{aligned} \tag{3.50}$$

$Z_i = \frac{\delta_i(g(T_i,\Lambda_0(T_i))+\tilde{h}(T_i))}{\frac{n-i+1}{n}}, \quad i = 1, 2, \ldots, n$ *and* $\hat{\theta} = \int \tilde{h}(t) d\Lambda_0(t)$.

In Theorem 3.2.2, if the jump of Λ *at* T_i *is (3.48), the remaining term is* $o_p\left(\frac{1}{\sqrt{n}}\right)$.

Proof Let $w_i = \Delta\Lambda(T_i)$ for $i = 1, \ldots, n$ where we notice $w_n = \delta_n$. We can rewrite the constraint (3.47) in the discrete format. The constraint (3.47) for any cumulative hazard that is dominated by the Nelson-Aalen estimator can be written as

$$\sum_{i=1}^{n-1} \delta_i \left(g(T_i, \Lambda_0(T_i)) + \tilde{h}(T_i)\right) w_i + \delta_n \left(g(T_n, \Lambda_0(T_n)) + \tilde{h}(T_n)\right) - \hat{\theta} = \theta \tag{3.51}$$

Similarly, AL of this cumulative hazard can be written as

$$AL = \prod_{i=1}^{n} (w_i)^{\delta_i} \exp\left\{-\sum_{j=1}^{i} w_j\right\} \tag{3.52}$$

And the $\log AL$ is

$$\begin{aligned}
\log AL &= \sum_{i=1}^{n} \delta_i \log w_i - \sum_{i=1}^{n} \sum_{j=1}^{i} w_j \\
&= \sum_{i=1}^{n} \delta_i \log w_i - \sum_{j=1}^{n} \sum_{i=j}^{n} w_j \\
&= \sum_{i=1}^{n} \delta_i \log w_i - \sum_{i=1}^{n} (n-i+1)\, w_i
\end{aligned} \tag{3.53}$$

In order to use the Lagrange multiplier method, we form the target function G as follows.

$$\begin{aligned}
G = &\sum_{i=1}^{n} \delta_i \log w_i - \sum_{i=1}^{n} (n-i+1)\, w_i \\
&+ n\lambda \left[\theta + \hat{\theta} - \sum_{i=1}^{n-1} \delta_i \left(g(T_i, \Lambda_0(T_i)) + \tilde{h}(T_i)\right) w_i - \delta_n \left(g(T_n, \Lambda_0(T_n)) + \tilde{h}(T_n)\right)\right]
\end{aligned} \tag{3.54}$$

Taking the derivative with respect to w_i, $i = 1, \ldots, n-1$ and equaling them to 0 yields

$$\frac{\partial G}{\partial w_i} = \frac{\delta_i}{w_i} - (n-i+1) - n\lambda\delta_i\left(g(T_i,\Lambda_0(T_i)) + \tilde{h}(T_i)\right) = 0, \quad i = 1,\ldots,n-1 \tag{3.55}$$

Then the jump of Λ at T_i is

$$\begin{aligned} w_i = \Delta\Lambda(T_i) &= \frac{\delta_i}{(n-i+1) + n\lambda\delta_i\left(g(T_i,\Lambda_0(T_i)) + \tilde{h}(T_i)\right)} \\ &= \frac{\delta_i}{n-i+1}\frac{1}{1+\lambda\frac{\delta_i\left(g(T_i,\Lambda_0(T_i))+\tilde{h}(T_i)\right)}{\frac{n-i+1}{n}}} \\ &= \Delta\hat{\Lambda}_{NA}(T_i)\frac{1}{1+\lambda Z_i} \end{aligned} \tag{3.56}$$

where $Z_i = \frac{\delta_i(g(T_i,\Lambda_0(T_i))+\tilde{h}(T_i))}{\frac{n-i+1}{n}}$, $\quad i = 1,2,\ldots,n$ and $\Delta\hat{\Lambda}_{NA}(T_i) = \frac{\delta_i}{n-i+1}$.

Since $Z_n = n\delta_n\left(g(T_n,\Lambda_0(T_n)) + \tilde{h}(T_n)\right)$ and $\delta_i^2 = \delta_i,\ i = 1,\ldots,n,$ plugging the $w_i,\ i = 1,\ldots,n-1$ and $w_n = \delta_n$ into (3.51) gives us the equation to solve for λ as below.

$$\sum_{i=1}^{n-1}\frac{n-i+1}{n}Z_i \times \Delta\hat{\Lambda}_{NA}(T_i)\frac{1}{1+\lambda Z_i} + \frac{1}{n}\delta_n Z_n - \hat{\theta} = \theta \tag{3.57}$$

It can be simplified as

$$\frac{1}{n}\sum_{i=1}^{n-1}\frac{\delta_i Z_i}{1+\lambda Z_i} + \frac{1}{n}\delta_n Z_n - \hat{\theta} = \theta \tag{3.58}$$

In Theorem 3.2.2, if the jump of Λ at T_i is (3.48), then $\Lambda(t)$ is as follows. We denote $\Lambda(t)$ as $\hat{\Lambda}_n(t)$.

$$\hat{\Lambda}_n(t) = \sum_{T_i \le t} w_i \tag{3.59}$$

where $w_i = \Delta\hat{\Lambda}_{NA}(T_i)\frac{1}{1+\lambda Z_i}$ and $Z_i = \frac{\delta_i\left(g(T_i,\Lambda_0(T_i))+\tilde{h}(T_i)\right)}{\frac{n-i+1}{n}}$.

So the difference between $\hat{\Lambda}_n(t)$ and $\hat{\Lambda}_{NA}(t)$ is as follows.

$$\begin{aligned} \left|\hat{\Lambda}_n(t) - \hat{\Lambda}_{NA}(t)\right| &= \left|\sum_{T_i \le t} \Delta\hat{\Lambda}_{NA}(T_i)\frac{1}{1+\lambda Z_i} - \sum_{T_i \le t} \Delta\hat{\Lambda}_{NA}(T_i)\right| \\ &= \left|\sum_{T_i \le t} \Delta\hat{\Lambda}_{NA}(T_i)\frac{\lambda Z_i}{1+\lambda Z_i}\right| \end{aligned} \tag{3.60}$$

In Appendix, we know that $\lambda = O_p\left(\frac{1}{\sqrt{n}}\right)$ and $\max_{1\le n}|Z_i| = o_p(\sqrt{n})$, then we have $\max_{1\le n}|\lambda Z_i| = o_p(1)$, so we may expand $\frac{1}{1+\lambda Z_i}$ as follows.

$$\frac{1}{1+\lambda Z_i} = 1 - \lambda Z_i + O_p\left(\lambda^2\right) Z_i^2 \tag{3.61}$$

Then

$$\begin{aligned} \left|\hat{\Lambda}_n(t) - \hat{\Lambda}_{NA}(t)\right| &= \left|\sum_{T_i \le t} \Delta\hat{\Lambda}_{NA}(T_i)\lambda Z_i\left(1 - \lambda Z_i + O_p(\lambda^2)Z_i^2\right)\right| \\ &\le |\lambda|\left|\sum_{T_i \le t} \Delta\hat{\Lambda}_{NA}(T_i)Z_i\right| + \lambda^2\left|\sum_{T_i \le t} \Delta\hat{\Lambda}_{NA}(T_i)Z_i^2\right| \\ &+ O_p\left(\lambda^3\right)\left|\sum_{T_i \le t} \Delta\hat{\Lambda}_{NA}(T_i)Z_i^3\right| \end{aligned} \tag{3.62}$$

Since we have

$$\begin{aligned} \left|\sum_{T_i \le t} \Delta\hat{\Lambda}_{NA}(T_i)Z_i\right| &= \left|\sum_{T_i \le t} \Delta\hat{\Lambda}_{NA}(T_i)\frac{\delta_i\left(g(T_i, \Lambda_0(T_i)) + \tilde{h}(T_i)\right)}{\frac{n-i+1}{n}}\right| \\ &= \left|\int_0^t \frac{g(s, \Lambda_0(s)) + \tilde{h}(s)}{\frac{Y(s)}{n}} d\hat{\Lambda}_{NA}(s)\right| \end{aligned} \tag{3.63}$$

where $Y(s) = \sum_{i=1}^n I[Y_i \ge s]$.

Similar to the arguments of Pan and Zhou (2002) Lemma A3, we have

$$\int_0^t \frac{g(s, \Lambda_0(s)) + \tilde{h}(s)}{\frac{Y(s)}{n}} d\hat{\Lambda}_{NA}(s) \xrightarrow{P} \int_0^t \frac{g(s, \Lambda_0(s)) + \tilde{h}(s)}{(1 - F_0(s))(1 - G_o(s))} d\Lambda_0(s) \tag{3.64}$$

so we have

$$\left|\sum_{T_i \le t} \Delta\hat{\Lambda}_{NA}(T_i) Z_i\right| = O_p(1) \tag{3.65}$$

Similarly, we also have

$$\left|\sum_{T_i \le t} \Delta\hat{\Lambda}_{NA}(T_i) Z_i^2\right| = O_p(1), \quad \left|\sum_{T_i \le t} \Delta\hat{\Lambda}_{NA}(T_i) Z_i^3\right| = O_p(1) \tag{3.66}$$

Since $|\lambda| = O_p(\frac{1}{\sqrt{n}})$, we have

$$\left|\hat{\Lambda}_n(t) - \hat{\Lambda}_{NA}(t)\right| \le O_p\left(\frac{1}{\sqrt{n}}\right), \quad for\ any\ t \in [0, \tau] \tag{3.67}$$

so

$$\left\|\hat{\Lambda}_n(t) - \hat{\Lambda}_{NA}(t)\right\| = \sup_{t \in [0,\tau]} \left|\hat{\Lambda}_n(t) - \hat{\Lambda}_{NA}(t)\right| = O_p\left(\frac{1}{\sqrt{n}}\right) \tag{3.68}$$

As is well known,

$$\left\|\hat{\Lambda}_{NA}(t) - \Lambda_0(t)\right\| = \sup_{t \in [0,\tau]} \left|\hat{\Lambda}_{NA}(t) - \Lambda_0(t)\right| = O_p\left(\frac{1}{\sqrt{n}}\right) \tag{3.69}$$

By the triangle inequality of the sup norm, we have

$$\left\|\hat{\Lambda}_n(t) - \Lambda_0(t)\right\| \le \left\|\hat{\Lambda}_n(t) - \hat{\Lambda}_{NA}(t)\right\| + \left\|\hat{\Lambda}_{NA}(t) - \Lambda_0(t)\right\| = O_p\left(\frac{1}{\sqrt{n}}\right) \tag{3.70}$$

Therefore, the remaining term is

$$\left|T(\hat{\Lambda}_n) - T(\Lambda_0) - T'_{\Lambda_0}\left(\hat{\Lambda}_n - \Lambda_0\right)\right| = o\left(\left\|\hat{\Lambda}_n - \Lambda_0\right\|\right) = o_p\left(\frac{1}{\sqrt{n}}\right) \tag{3.71}$$

Next, we would like to have a discussion of the feasibility and the feasible values of θ.

The function $l(\lambda)$ defined in (3.50) is monotone decreasing and continuous with respect to λ, which can be verified by taking the first derivative with respect to λ. Any legitimate value λ must result in w_i bounded between zero and one. This restriction leads to the following legitimate λ range Φ.

All max and min in the following definition are taken in the domain $\{i : 1 \leq i \leq n-1, \delta_i = 1, \ and \ g(T_i) \neq 0\}$. If there are any additional restrictions, we will specify in each individual case.

Case 1. When $\min\left(g(T_i, \Lambda_0(T_i)) + \tilde{h}(T_i)\right) > 0$

$$\Phi = \left(\max \frac{i-n}{n\left(g(T_i, \Lambda_0(T_i)) + \tilde{h}(T_i)\right)}, \infty\right) := (\underline{\lambda}, \infty) \tag{3.72}$$

Case 2. When $\max\left(g(T_i, \Lambda_0(T_i)) + \tilde{h}(T_i)\right) < 0$

$$\Phi = \left(-\infty, \min \frac{i-n}{n\left(g(T_i, \Lambda_0(T_i)) + \tilde{h}(T_i)\right)}\right) := \left(-\infty, \bar{\lambda}\right) \tag{3.73}$$

Case 3. When $\max\left(g(T_i, \Lambda_0(T_i)) + \tilde{h}(T_i)\right) > 0 > \min\left(g(T_i, \Lambda_0(T_i)) + \tilde{h}(T_i)\right)$

$$\Phi = \left(\max \frac{i-n}{n\left(g(T_i, \Lambda_0(T_i)) + \tilde{h}(T_i)\right)}, \min \frac{i-n}{n\left(g(T_i, \Lambda_0(T_i)) + \tilde{h}(T_i)\right)}\right) := (\underline{\lambda}, \bar{\lambda}) \tag{3.74}$$

Because the function $l(\lambda)$ is monotone and continuous, the corresponding range of the θ value that makes (3.47) feasible is as follows. Define $\tilde{G}(t) = g(t, \Lambda_0(t)) + \tilde{h}(t)$

Case 1.

$$\mathscr{V} = \left(\tilde{G}(T_n)\delta_n, \sum_{i=1}^{n-1} \frac{\delta_i \tilde{G}(T_i)}{n-i+1+n\underline{\lambda}\tilde{G}(T_i)} + \tilde{G}(T_n)\delta_n \right) \tag{3.75}$$

Case 2.

$$\mathscr{V} = \left(\sum_{i=1}^{n-1} \frac{\delta_i \tilde{G}(T_i)}{n-i+1+n\bar{\lambda}\tilde{G}(T_i)} + \tilde{G}(T_n)\delta_n, \tilde{G}(T_n)\delta_n \right) \tag{3.76}$$

Case 3.

$$\mathscr{V} = \left(\sum_{i=1}^{n-1} \frac{\delta_i \tilde{G}(T_i)}{n-i+1+n\bar{\lambda}\tilde{G}(T_i)} + \tilde{G}(T_n)\delta_n, \sum_{i=1}^{n-1} \frac{\delta_i \tilde{G}(T_i)}{n-i+1+n\underline{\lambda}\tilde{G}(T_i)} + \tilde{G}(T_n)\delta_n \right) \tag{3.77}$$

Next lemma shows that the limiting distribution of $n\lambda^2$ is a χ_1^2 distribution times a constant.

Lemma 3.2.4 *Suppose $g(t, \Lambda(t))$ is left continuous with respect to t and twice differentiable with respect to $\Lambda(t)$ and satisfies*

$$0 < \int \frac{\left(g(t, \Lambda_0(t)) + \tilde{h}(t)\right)^2}{(1-F_0(t))(1-G_0(t))} d\Lambda_0(t) < \infty \tag{3.78}$$

where $\tilde{h}(t) = \int_t^{+\infty} h(s, \Lambda_0(s)) d\Lambda_0(s)$ and $h = \frac{\partial g}{\partial \Lambda}$.
Then $\theta_0 = \int g(t, \Lambda_0(t)) d\Lambda_0(t)$ is feasible with probability approaching 1 as $n \to \infty$, and the solution λ of (3.49) with $\theta = \theta_0$ satisfies

$$n\lambda^2 \xrightarrow{D} \chi^2_{(1)} \left(\int \frac{\left(g(t, \Lambda_0(t)) + \tilde{h}(t)\right)^2}{(1-F_0(t))(1-G_0(t))} d\Lambda_0(t) \right)^{-1}, \ as\ n \to \infty \tag{3.79}$$

Proof See Appendix. ∎

Lemma 3.2.5 *Let* $(T_1, \delta_1), \ldots, (T_n, \delta_n)$ *be* n *pairs of random variables defined in (3.42). Suppose* $g(t, \Lambda(t))$ *is left continuous with respect to* t *and twice differentiable with respect to* Λ *and satisfies*

$$0 < \int \frac{\Big(g(t, \Lambda_0(t)) + \tilde{h}(t)\Big)^2}{(1 - F_0(t))(1 - G_0(t))} d\Lambda_0(t) < \infty \tag{3.80}$$

ALR is defined by

$$ALR(\theta) = \frac{sup\Big\{AL(\Lambda)|\Lambda \ll \hat{\Lambda}_{NA}, and \quad \Lambda \quad satisfy \quad (3.47)\Big\}}{AL(\hat{\Lambda}_{NA})} \tag{3.81}$$

Then

$$-2logALR(\theta_0) \xrightarrow{D} \chi^2_{(1)} \qquad as \quad n \to \infty \tag{3.82}$$

Proof Since

$$AL(\Lambda) = \prod_{i=1}^{n} \left[\Delta\Lambda\left(T_i\right)\right]^{\delta_i} \exp\left\{-\Lambda\left(T_i\right)\right\} \tag{3.83}$$

Denote $\Delta\Lambda\left(T_i\right) = w_i$

$$\begin{aligned}
\log AL(\Lambda) &= \sum_{i=1}^{n}\left(\delta_i \log w_i - \sum_{j=1}^{i} w_j\right) = \sum_{i=1}^{n} \delta_i \log w_i - \sum_{i=1}^{n}\sum_{j=1}^{i} w_j \\
&= \sum_{i=1}^{n} \delta_i \log w_i - \sum_{j=1}^{n}\sum_{i=j}^{n} w_j = \sum_{i=1}^{n} \delta_i \log w_i - \sum_{j=1}^{n} (n - j + 1)\, w_j \\
&= \sum_{i=1}^{n} \delta_i \log w_i - \sum_{i=1}^{n} (n - i + 1)\, w_i
\end{aligned} \tag{3.84}$$

Define $Z_i = \frac{\delta_i\left(g(T_i, \Lambda_0(T_i)) + \tilde{h}(T_i)\right)}{\frac{n-i+1}{n}}$, $i = 1, \ldots, n.$

Consider

$$
\begin{aligned}
& -2\log ALR(\theta_0) = 2\left[\sum_{i=1}^{n}\delta_i \log \Delta\hat{\Lambda}_{NA}(T_i) - \sum_{i=1}^{n}(n-i+1)\,\Delta\hat{\Lambda}_{NA}(T_i)\right] \\
& -2\left[\sum_{i=1}^{n-1}\delta_i \log\left(\Delta\hat{\Lambda}_{NA}(T_i)\frac{1}{1+\lambda Z_i}\right) - \sum_{i=1}^{n-1}(n-i+1)\,\Delta\hat{\Lambda}_{NA}(T_i)\frac{1}{1+\lambda Z_i}\right] \\
& -2\left[\delta_n \log \Delta\hat{\Lambda}_{NA}(T_n) - \Delta\hat{\Lambda}_{NA}(T_n)\right] \\
& = 2\sum_{i=1}^{n-1}\delta_i \log(1+\lambda Z_i) - 2\sum_{i=1}^{n-1}(n-i+1)\,\Delta\hat{\Lambda}_{NA}(T_i)\frac{\lambda Z_i}{1+\lambda Z_i} \\
& = 2\sum_{i=1}^{n-1}\delta_i \log(1+\lambda Z_i) - 2\sum_{i=1}^{n-1}\frac{\delta_i \lambda Z_i}{1+\lambda Z_i} \\
& = 2\sum_{i=1}^{n-1}\delta_i \log(1+\lambda Z_i) - 2\sum_{i=1}^{n-1}\delta_i \lambda Z_i + 2\sum_{i=1}^{n-1}\frac{\delta_i \lambda^2 Z_i^2}{1+\lambda Z_i}
\end{aligned}
\tag{3.85}
$$

In Appendix, note that

$\max_{1\leq i\leq n}|\lambda Z_i| = |\lambda|\max_{1\leq i\leq n}|Z_i| = O_p\left(\frac{1}{\sqrt{n}}\right)o_p(\sqrt{n}) = o_p(1)$, we may expand $\log(1+\lambda Z_i)$ as follows.

$$
\log(1+\lambda Z_i) = \lambda Z_i - \frac{1}{2}\lambda^2 Z_i^2 + O_p\left(\lambda^3\right)Z_i^3 \tag{3.86}
$$

Substituting this into the expression of $-2\log ALR(\theta_0)$ gives us,

$$
\begin{aligned}
-2\log ALR(\theta_0) &= -\sum_{i=1}^{n-1}\delta_i \lambda_i^2 Z_i^2 + 2O_p\left(\lambda^3\right)\sum_{i=1}^{n-1}Z_i^3 + 2\sum_{i=1}^{n-1}\frac{\delta_i \lambda^2 Z_i^2}{1+\lambda Z_i} \\
&= -\sum_{i=1}^{n-1}\delta_i \lambda_i^2 Z_i^2 + 2O_p\left(\lambda^3\right)\sum_{i=1}^{n-1}Z_i^3 + 2\sum_{i=1}^{n-1}\delta_i \lambda^2 Z_i^2 - 2\sum_{i=1}^{n-1}\frac{\delta_i \lambda^3 Z_i^3}{1+\lambda Z_i} \\
&= \sum_{i=1}^{n-1}\delta_i \lambda_i^2 Z_i^2 + 2O_p\left(\lambda^3\right)\sum_{i=1}^{n-1}Z_i^3 - 2\sum_{i=1}^{n-1}\frac{\delta_i \lambda^3 Z_i^3}{1+\lambda Z_i} \\
&= n\lambda^2\frac{1}{n}\sum_{i=1}^{n-1}\delta_i Z_i^2 + 2O_p\left(\lambda^3\right)\sum_{i=1}^{n-1}Z_i^3 - 2\sum_{i=1}^{n-1}\frac{\delta_i \lambda^3 Z_i^3}{1+\lambda Z_i}
\end{aligned}
\tag{3.87}
$$

Since we have

$$\begin{aligned} &\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n-1}\delta_i Z_i^2 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n-1} Z_i^2 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} Z_i^2 \\ &= \int \frac{\left(g(t,\Lambda_0(t)) + \tilde{h}(t)\right)^2}{(1-F(t))\,(1-G(t))} d\Lambda_0(t) < \infty \end{aligned} \tag{3.88}$$

where the limitation is under the meaning of converging in probability, and the following terms are negligible.

$$\begin{aligned} &\left| O_p\left(\lambda^3\right)\sum_{i=1}^{n-1} Z_i^3 \right| \le O_p\left(n^{-\frac{1}{2}}\right) o_p\left(n^{\frac{1}{2}}\right)\frac{1}{n}\sum_{i=1}^{n} Z_i^2 = o_p(1) \\ &\sum_{i=1}^{n-1}\frac{\delta_i\lambda^3 Z_i^3}{1+\lambda Z_i} \le O_p\left(n^{-\frac{1}{2}}\right) o_p\left(n^{\frac{1}{2}}\right)\frac{1}{n}\sum_{i=1}^{n} Z_i^2 = o_p(1) \end{aligned} \tag{3.89}$$

and we also have,

$$n\lambda^2 \xrightarrow{D} \chi^2_{(1)} \left(\int \frac{\left(g(t,\Lambda_0(t)) + \tilde{h}(t)\right)^2}{(1-F_0(t))(1-G_0(t))} d\Lambda_0(t) \right)^{-1} \tag{3.90}$$

By the Slutsky theorem, we have

$$-2\log ALR(\theta_0) \xrightarrow{D} \chi^2_{(1)} \quad as \quad n\to\infty \tag{3.91}$$

∎

In the following theorem, we prove that the difference between EL and AL is negligible when n is large in the case of discrete cumulative hazard functions. And we prove that

$$-2\log ELR(\theta_0) \xrightarrow{D} \chi^2_{(1)} \; as \; n\to\infty \tag{3.92}$$

under the null hypothesis as follows

$$H_0 : T(\Lambda_0) + T'_{\Lambda_0}(\Lambda - \Lambda_0) = \theta_0 \tag{3.93}$$

where $ELR(\theta)$ function is defined in the following theorem.

Theorem 3.2.6 *Suppose all conditions of Lemma (3.2.5) hold and ELR is defined by*

$$ELR(\theta) = \frac{EL(\Lambda^*)}{EL(\hat{\Lambda}_{NA})} \tag{3.94}$$

where Λ^* is given by the jumps defined in (3.48)

Then we have,

$$-2logELR(\theta_0) \xrightarrow{D} \chi^2_{(1)}, \quad as \quad n \to \infty \tag{3.95}$$

Proof Remember that

$$EL(\Lambda) = \prod_{i=1}^{n} \left\{ [\Delta\Lambda(T_i)]^{\delta_i} \left[\prod_{j:T_j<T_i} (1 - \Delta\Lambda(T_j)) \right]^{\delta_i} \left[\prod_{j:T_j\leq T_i} (1 - \Delta\Lambda(T_j)) \right]^{1-\delta_i} \right\} \tag{3.96}$$

Denote $\Delta\Lambda(T_i) = w_i$, $\log EL(\Lambda)$ can be written as follows.

$$\begin{aligned}
\log EL(\Lambda) &= \sum_{i=1}^{n} \delta_i \log w_i + \sum_{i=1}^{n} \delta_i \sum_{j=1}^{i-1} \log(1 - w_j) + \sum_{i=1}^{n} (1 - \delta_i) \sum_{j=1}^{i} \log(1 - w_j) \\
&= \sum_{i=1}^{n} \delta_i \log w_i + \sum_{i=1}^{n} \sum_{j=1}^{i} \log(1 - w_j) - \sum_{i=1}^{n} \delta_i \log(1 - w_i) \\
&= \sum_{i=1}^{n} \delta_i \log w_i + \sum_{j=1}^{n} \sum_{i=j}^{n} \log(1 - w_j) - \sum_{i=1}^{n} \delta_i \log(1 - w_i) \\
&= \sum_{i=1}^{n} \delta_i \log w_i + \sum_{i=1}^{n} (n - i + 1) \log(1 - w_i) - \sum_{i=1}^{n} \delta_i \log(1 - w_i) \\
&= \sum_{i=1}^{n} \delta_i \log w_i + \sum_{i=1}^{n} (n - i + 1 - \delta_i) \log(1 - w_i)
\end{aligned} \tag{3.97}$$

Then we have

$$\begin{aligned}
&-2\log ELR(\theta_0)\\
&=2\left[\sum_{i=1}^{n}\delta_i \log \Delta\hat{\Lambda}_{NA}(T_i)+\sum_{i=1}^{n}(n-i+1-\delta_i)\log\Big(1-\Delta\hat{\Lambda}_{NA}(T_i)\Big)\right]\\
&-2\sum_{i=1}^{n-1}\delta_i\log\left(\Delta\hat{\Lambda}_{NA}(T_i)\frac{1}{1+\lambda Z_i}\right)\\
&-2\sum_{i=1}^{n-1}(n-i+1-\delta_i)\log\left(1-\Delta\hat{\Lambda}_{NA}(T_i)\frac{1}{1+\lambda Z_i}\right) \qquad (3.98)\\
&-2\Big[\delta_n\log\Delta\hat{\Lambda}_{NA}(T_n)+(1-\delta_n)\log\Big(1-\Delta\hat{\Lambda}_{NA}(T_n)\Big)\Big]\\
&=2\sum_{i=1}^{n-1}\delta_i\log(1+\lambda Z_i)+2\sum_{i=1}^{n-1}(n-i+1-\delta_i)\log\Big(1-\Delta\hat{\Lambda}_{NA}(T_i)\Big)\\
&-2\sum_{i=1}^{n-1}(n-i+1-\delta_i)\log\left(1-\Delta\hat{\Lambda}_{NA}(T_i)\frac{1}{1+\lambda Z_i}\right)
\end{aligned}$$

For Taylor expansion with Lagrange remaining term, when $|x-x_0|=o(1)$, we have

$$\log x=\log x_0+\frac{1}{x_0}(x-x_0)+\frac{1}{2}\left(-\frac{1}{x_0^2}\right)\eta^2 \qquad (3.99)$$

where $|\eta|\le|x-x_0|$.

Since we have

$$\begin{aligned}
&\left(1-\Delta\hat{\Lambda}_{NA}(T_i)\frac{1}{1+\lambda Z_i}\right)=\log\left(1-\Delta\hat{\Lambda}_{NA}(T_i)+\Delta\hat{\Lambda}_{NA}(T_i)\frac{\lambda Z_i}{1+\lambda Z_i}\right)\\
& \qquad (3.100)\\
&\left|\Delta\hat{\Lambda}_{NA}(T_i)\frac{\lambda Z_i}{1+\lambda Z_i}\right|=o_p(1)
\end{aligned}$$

choose $x = 1 - \Delta\hat{\Lambda}_{NA}(T_i)\frac{1}{1+\lambda Z_i}$ and $x_0 = 1 - \Delta\hat{\Lambda}_{NA}(T_i)$, we have

$$\begin{aligned}
&\log\left(1 - \Delta\hat{\Lambda}(T_i)\frac{1}{1+\lambda Z_i}\right) = \log\Big(1 - \Delta\hat{\Lambda}_{NA}(T_i)\Big) \\
&+ \frac{1}{1 - \Delta\hat{\Lambda}_{NA}(T_i)}\Delta\hat{\Lambda}_{NA}(T_i)\frac{\lambda Z_i}{1+\lambda Z_i} \\
&+ \frac{1}{2}\left(-\frac{1}{\Big(1 - \Delta\hat{\Lambda}_{NA}(T_i)\Big)^2}\right)\Big(\Delta\hat{\Lambda}_{NA}(T_i)\Big)^2 \eta_i^2 \\
&= \log\Big(1 - \Delta\hat{\Lambda}_{NA}(T_i)\Big) + \frac{\Delta\hat{\Lambda}_{NA}(T_i)}{1 - \Delta\hat{\Lambda}_{NA}(T_i)}\frac{\lambda Z_i}{1+\lambda Z_i} - \frac{1}{2}\left(\frac{\Delta\hat{\Lambda}_{NA}(T_i)}{1 - \Delta\hat{\Lambda}_{NA}(T_i)}\right)^2 \eta_i^2
\end{aligned} \tag{3.101}$$

where $|\eta_i| \le \left|\frac{\lambda Z_i}{1+\lambda Z_i}\right|$. We notice that

$$\frac{\Delta\hat{\Lambda}_{NA}(T_i)}{1 - \Delta\hat{\Lambda}_{NA}(T_i)} = \frac{\frac{\delta_i}{n-i+1}}{1 - \frac{\delta_i}{n-i+1}} = \frac{\delta_i}{n-i+1-\delta_i} \tag{3.102}$$

Then 3.101 can be simplified as

$$\begin{aligned}
\log\left(1 - \Delta\hat{\Lambda}(T_i)\frac{1}{1+\lambda Z_i}\right) &= \log\Big(1 - \Delta\hat{\Lambda}_{NA}(T_i)\Big) + \frac{1}{(n-i+1-\delta_i)}\frac{\delta_i\lambda Z_i}{1+\lambda Z_i} \\
&- \frac{1}{2}\frac{\delta_i\eta_i^2}{(n-i+1-\delta_i)^2}
\end{aligned} \tag{3.103}$$

so

$$-2\log ELR(\theta_0) = 2\sum_{i=1}^{n-1}\delta_i\log(1+\lambda Z_i) - 2\sum_{i=1}^{n-1}\frac{\delta_i\lambda Z_i}{1+\lambda Z_i} + 2\sum_{i=1}^{n-1}\frac{\delta_i\eta_i^2}{(n-i+1-\delta_i)} \tag{3.104}$$

Remember that

$$-2\log ALR(\theta_0) = 2\sum_{i=1}^{n-1}\delta_i\log(1+\lambda Z_i) - 2\sum_{i=1}^{n-1}\frac{\delta_i\lambda Z_i}{1+\lambda Z_i} \tag{3.105}$$

so

$$-2\log ELR(\theta_0) + 2\log ALR(\theta_0) = 2\sum_{i=1}^{n-1} \frac{\delta_i \eta_i^2}{n-i+1-\delta_i} \tag{3.106}$$

In Appendix, we have

$$\frac{1}{n}\sum_{i=1}^{n} \frac{Z_i^2}{n-i} \xrightarrow{P} 0, \quad as \quad n \to \infty \tag{3.107}$$

so

$$\frac{1}{n}\sum_{i=1}^{n} \frac{Z_i^2}{n-i} = o_p(1) \tag{3.108}$$

By the following inequation,

$$\begin{aligned} 0 \le \sum_{i=1}^{n-1} \frac{\delta_i \eta_i^2}{n-i+1-\delta_i} &\le \lambda^2 \sum_{i=1}^{n} \frac{Z_i^2}{n-i+1-\delta_i} = n\lambda^2 \frac{1}{n}\sum_{i=1}^{n} \frac{Z_i^2}{n-i+1-\delta_i} \\ &\le n\lambda^2 \frac{1}{n}\sum_{i=1}^{n} \frac{Z_i^2}{n-i} = O_p(1)o_p(1) = o_p(1) \end{aligned} \tag{3.109}$$

We have

$$-2\log ELR(\theta_0) + 2\log ALR(\theta_0) \xrightarrow{P} 0 \tag{3.110}$$

By the Slutsky theorem, we have

$$-2\log ELR(\theta_0) \xrightarrow{D} \chi^2_{(1)} \quad as \quad n \to \infty \tag{3.111}$$

∎

In the following theorem, we prove that

$$-2\log ELR(\theta_0) \xrightarrow{D} \chi^2_{(1)} \; as \; n \to \infty \tag{3.112}$$

under the null hypothesis

$$H_0 : \int g(t, \Lambda(t)) d\Lambda(t) = \theta_0 \tag{3.113}$$

where ELR function is defined in the following theorem.

Theorem 3.2.7 *Suppose the non-linear statistical functional* $T(\Lambda) = \int g(t, \Lambda(t)) d\Lambda(t)$ *is Hadamard differentiable at* $\Lambda_0(t)$*. The sufficient conditions for Hadamard differentiable are listed in Theorem 3.2.2. Empirical likelihood ratio (ELR) function is defined as follows.*

$$ELR = \frac{EL(\Lambda)}{EL(\hat{\Lambda}_{NA})} \tag{3.114}$$

where $EL(\cdot)$ *is defined in 2.15 and* $\hat{\Lambda}_{NA}$ *is the Nelson-Aalen estimator.*
If the following null hypothesis is true

$$H_0 : \int g(t, \Lambda(t)) d\Lambda(t) = \theta_0 \tag{3.115}$$

where $\theta_0 = \int g(t, \Lambda_0(t)) d\Lambda_0(t)$*.*
then we have

$$-2 \log ELR(\theta_0) \xrightarrow{D} \chi^2_{(1)} \tag{3.116}$$

where $ELR(\theta_0)$ *is the maximum of the ELR function (3.114) subject to the non-linear statistical functional constraint* $\int g(t, \Lambda(t)) d\Lambda(t) = \theta_0$*.*

Proof This theorem is a straightforward result of Theorem 3.2.1, Theorem 3.2.2 and Theorem 3.2.6. ∎

Chapter 4 Empirical Likelihood Ratio Subject to Nonlinear Statistical Functional in Terms of Distribution Function with Uncensored Data

4.1 Introduction

Background

This chapter deals with uncensored data. Owen(1988) proves that $-2\log ELR$ converges to a χ_1^2 subject to a linear statistical functional in terms of distribution functions with uncensored data. He also mentions how to deal with the non-linear but Frechet differentiable statistical functional constraint in terms of distribution functions. However, he does not specify the form of the constraint. Moreover, Frechet differentiability is too strong and a lot of statistics are not Frechet differentiable (See Example 2.3.2 Fernholz (1983)). In this chapter, we prove that under some regularity conditions the non-linear statistical functional $T(F) = \int g(t, F(t))dF(t)$ is Hadamard differentiable. Hadamard differentiability is a weaker form of differentiability which can be applied to a large class of statistics. On the other hand, it also allows the functional delta method to carry through. Gateaux differentiability is weaker than Hadamard differentiability. However, it needs extra conditions such as second order derivative to make the functional delta method work. The sufficient conditions for the Hadamard differentiability are listed in Theorem 4.2.3. And we prove that

$$-2\log ELR(\theta_0) \xrightarrow{D} \chi^2_{(1)} \; as \; n \to \infty \tag{4.1}$$

under the null hypothesis

$$H_0 : \int g(t, F(t))dF(t) = \theta_0 \tag{4.2}$$

where $ELR(\theta_0)$ is the maximum of the ELR function, which is defined later in (4.3), subject to the non-linear constraint $\int g(t, F(t))dF(t) = \theta_0$. The empirical likelihood ratio (ELR) function is defined as follows.

$$ELR = \frac{EL(F)}{EF(\hat{F}_n)} \tag{4.3}$$

where $EL(\cdot)$ is defined in (2.3) and $\hat{F}_n$ is the empirical distribution function.

Motivation

Our motivation to analyze this kind of constraint is that the constraint of the Gini index can be represented in the following form

$$\int g(t, F(t))dF(t) = \theta_0 \tag{4.4}$$

In Chapter 3, we mentioned that the Gini index (G) is just half of the Gini's relative mean difference as follows.

$$G = \frac{\int_0^{+\infty} y\left(2F(y) - 1\right) dF(y)}{\mu} \tag{4.5}$$

where $\mu = \int_0^{+\infty} ydF(y)$.

Suppose we would like to do the hypothesis testing of the Gini index. The null hypothesis is

$$G = \mu_0 \tag{4.6}$$

By simple transformations, (4.6) is equivalent to

$$\int \left(2yF(y) - (1 + \mu_0)\, y\right) dF(y) = 0 \tag{4.7}$$

which is the form of (4.4).

4.2 Lemma and Theorem

First of all, we propose a theorem similar to Theorem 3.2.1.

Theorem 4.2.1 *Suppose we have two statistical functional constraints in terms of distribution functions as follows.*

$$T_1(F) = \theta_0, \quad T_2(F) = \theta_0 \tag{4.8}$$

which satisfy $\theta_0 = T_1(F_0)$ and $\theta_0 = T_2(F_0)$.
If for any statistic $\hat{F}_n(t)$ (not necessarily empirical distribution function), s.t. $\left\|\hat{F}_n(t) - F_0(t)\right\| = O_p\left(\frac{1}{\sqrt{n}}\right)$, we have

$$\left|T_1(\hat{F}_n) - T_2(\hat{F}_n)\right| = o_p\left(\frac{1}{\sqrt{n}}\right) \tag{4.9}$$

then we have

$$-2\log ELR_1(\theta_0) + 2\log ELR_2(\theta_0) \xrightarrow{P} 0, \text{ as } n \to \infty \tag{4.10}$$

where $ELR_1(\theta_0)$ and $ELR_2(\theta_0)$ are the maximums of the ELR function (4.3) subject to the corresponding constraint $T_1(F) = \theta_0$ and $T_2(F) = \theta_0$ respectively.

Proof Similar to the proof of Theorem 3.2.1. See Pan and Zhou (2002) for a detailed discussion. ∎

In the following lemma, we calculate the influence curve of $T(F) = \int g(t, F(t))dF(t)$.

Lemma 4.2.2 *Suppose $g(t, F(t))$ is left continuous with respect to t and twice differentiable with respect to F. Let T be a statistical functional defined on the set of all* *distribution functions and* $T(F) = \int g(t, F(t))dF(t)$. *Suppose* $|g(t, F(t))| \leq G(t)$ *and* $|h(t, F(t))| \leq H(t)$, *for* $\forall$ t *and* F, *where* $G(t)$ *and* $H(t)$ *are integrable with respect to* t. *The influence curve of* T *at* F_0 *is*

$$
\begin{aligned}
IC_{T,F_0}(x) = g(x, F_0(x)) + \int_x^\infty h(t, F_0(t))dF_0(t) - \int h(t, F_0(t))F_0(t)dF_0(t) \\
- \int g(t, F_0(t))dF_0(t)
\end{aligned}
\tag{4.11}
$$

where, $h(t, F(t)) = \frac{\partial g}{\partial F}$.

Proof By the definition, the influence curve of $T(F) = \int g(t, F(t))dF(t)$ is

$$
\begin{aligned}
&IC_{T,F_0}(x) = \lim_{\epsilon \downarrow 0} \frac{\{T\left[(1-\epsilon)F_0 + \epsilon\delta_x\right] - T(F_0)\}}{\epsilon} \\
&= \lim_{\epsilon \downarrow 0} \frac{\int g(t, F_0 + \epsilon(\delta_x - F_0))d(F_0 + \epsilon(\delta_x - F_0)) - \int g(t, F_0)dF_0}{\epsilon} \\
&= \lim_{\epsilon \downarrow 0} \int \frac{(g(t, F_0 + \epsilon(\delta_x - F_0)) - g(t, F_0))}{\epsilon} dF_0 + \lim_{\epsilon \downarrow 0} \int g(t, F_0 + \epsilon(\delta_x - F_0))d(\delta_x - F_0)
\end{aligned}
\tag{4.12}
$$

By the L'Hospital's rule, we have

$$
\begin{aligned}
&\lim_{\epsilon \downarrow 0} \frac{g\left(t, F_0 + \epsilon\left(\delta_x - F_0\right)\right) - g(t, F_0)}{\epsilon} = \lim_{\epsilon \downarrow 0} \frac{\frac{d(g(t, F_0 + \epsilon(\delta_x - F_0)) - g(t, F_0))}{d\epsilon}}{\frac{d\epsilon}{d\epsilon}} \\
&= \lim_{\epsilon \downarrow 0} h\left(t, F_0 + \epsilon\left(\delta_x - F_0\right)\right)\left(\delta_x - F_0\right) = h\left(t, F_0\right)\left(\delta_x - F_0\right)
\end{aligned}
\tag{4.13}
$$

Since $g(t, F(t))$ is twice differentiable with respect to F, it is easy to verify that $g\left(t, F_0 + x\left(\delta_x - F_0\right)\right)$ is continuous with respect to x in $[0, \epsilon]$ and differentiable with respect to x in $(0, \epsilon)$. The derivative of $g\left(t, F_0 + x\left(\delta_x - F_0\right)\right)$ with respect to x is $h\left(t, F_0 + x\left(\delta_x - F_0\right)\right)\left(\delta_x - F_0\right)$. By **Mean Value Theorem**, there exists a $\xi \in [0, \epsilon]$, such that

$$
\frac{g\left(t, F_0 + \epsilon\left(\delta_x - F_0\right)\right) - g(t, F_0)}{\epsilon} = h\left(t, F_0 + \xi\left(\delta_x - F_0\right)\right)\left(\delta_x - F_0\right) \tag{4.14}
$$

Since $|h(t,F)| \leq H(t)$ and $\delta_x - F_0$ is uniformly bounded, $h\left(t, F_0 + \xi\left(\delta_x - F_0\right)\right)\left(\delta_x - F_0\right)$ is uniformly bounded as well.

Therefore, by Dominated Convergence Theorem,

$$\begin{aligned}
&\lim_{\epsilon \downarrow 0} \int \frac{(g(t, F_0 + \epsilon(\delta_x - F_0)) - g(t, F_0))}{\epsilon} dF_0 \\
&= \int \lim_{\epsilon \downarrow 0} \frac{(g(t, F_0 + \epsilon(\delta_x - F_0)) - g(t, F_0))}{\epsilon} dF_0 \\
&= \int h\left(t, F_0\right)\left(\delta_x - F_0\right) dF_0
\end{aligned} \tag{4.15}$$

Since $|g(t, F_0 + \epsilon(\delta_x - F_0))| \leq G(t)$, again, by Dominated Convergence Theorem, we have

$$\begin{aligned}
&\lim_{\epsilon \downarrow 0} \int g(t, F_0 + \epsilon(\delta_x - F_0)) d(\delta_x - F_0) = \int \lim_{\epsilon \downarrow 0} g(t, F_0 + \epsilon(\delta_x - F_0)) d(\delta_x - F_0) \\
&= \int g\left(t, F_0\right) d\left(\delta_x - F_0\right)
\end{aligned} \tag{4.16}$$

so (4.12) is

$$\begin{aligned}
&IC_{T,F_0}(x) = \int h(t, F_0)(\delta_x - F_0) dF_0 + \int g(t, F_0) d(\delta_x - F_0) \\
&= g(x, F_0(x)) + \int_x^{\infty} h(t, F_0(t)) dF_0(t) \\
&- \int h(t, F_0(t)) F_0(t) dF_0(t) - \int g(t, F_0(t)) dF_0(t) \\
&= g(x, F_0(x)) + \tilde{h}(x) - \int \tilde{h}(t) dF_0(t) - \int g(t, F_0(t)) dF_0(t)
\end{aligned} \tag{4.17}$$

where $\tilde{h}(x) = \int_x^{\infty} h(t, F_0(t)) dF_0(t)$.

The third part of the last equation is because

$$
\begin{aligned}
&\int \tilde{h}(x)dF_0(x) = \int \int_x^{\infty} h(t, F_0(t))dF_0(t)dF_0(x) \\
&= \int \left(\int_0^t dF_0(x) \right) h(t, F_0(t))dF_0(t) \\
&= \int h(t, F_0(t))F_0(t)dF_0(t)
\end{aligned}
\tag{4.18}
$$

∎

Note that the influence curve is just the Gateaux derivative in the direction of δ_x. And the existence of the influence curve for a statistical functional does not imply that the functional is Gateaux differentiable (See Example 2.2.2 of Fernholz (1983)). However, if T is Gateaux differentiable at F_0, the Gateaux derivative of T at F_0 may be written as

$$
\begin{aligned}
T'_{F_0}(F - F_0) &= \int IC_{T,F_0}(x)d(F(x) - F_0(x)) \\
&= \int \left(g(x, F_0(x)) + \tilde{h}(x)\right) d(F(x) - F_0(x))
\end{aligned}
\tag{4.19}
$$

Moreover, if T is Hadamard differentiable at F_0, the Hadamard derivative of T at F_0 is just the Gateaux derivative above.

In the following theorem, we prove that under some regularity conditions, T is Hadamard differentiable at F_0 with derivative T'_{F_0}.

Theorem 4.2.3 *Let $X_1, \ldots, X_n$ be n i.i.d. random variables with distribution function F_0, which has finite mean. Let $\mathbb{D}$ be the linear space expanded by $\mathbb{F}$. $\mathbb{F}$ is the set of all distribution functions with finite mean. $\mathbb{D}$ is equipped with the sup norm.*

$$
\|f\| = \sup_{x \in \mathbb{R}} |f(x)| \tag{4.20}
$$

Let T be a statistical functional defined on the $\mathbb{D}$ and $T(F) = \int g(t, F(t))dF(t)$. Define $h = \frac{\partial g}{\partial F}$ and $\tilde{h}(t) = \int_t^{+\infty} h(s, F_0(s))dF_0(s)$.

Under some regularity conditions,

Assumption (A) *$g(t, F(t))$ is left continuous with respect to t and twice differentiable with respect to $F(t)$.*

Assumption (B) *$|g(t, F(t))| \leq A(t)$, $|h(t, F(t))| \leq B(t)$ for all $t \in \mathbb{R}$ and $F(t) \in \mathbb{D}$, where $A(t)$ is integrable with respect to any distribution function $F(t) \in \mathbb{F}$ and $B(t)$ is integrable with respect to $F_0(t)$.*

T is Hadamard differentiable with derivative

$$T'_{F_0}(F - F_0) = \int \left(g(x, F_0(x)) + \tilde{h}(x)\right) d\left(F(x) - F_0(x)\right) \tag{4.21}$$

where $\tilde{h}(x) = \int_x^{\infty} h(t, F_0(t))dF_0(t)$ and $h = \frac{\partial g}{\partial F}$.

And the remaining term is

$$\left|T(F) - T(F_0) - T'_{F_0}(F - F_0)\right| = o\left(\|F(t) - F_0(t)\|\right) \tag{4.22}$$

In particular, if $F(t) = \hat{F}_n(t)$, where $\hat{F}_n(t)$ is the empirical distribution function, the remaining term is $o_p\left(\frac{1}{\sqrt{n}}\right)$

Proof In order to prove that T is Hadamard differentiable, we need to prove that for any $D_1, D_2, \ldots,$ and $D \in \mathbb{D}$, such that $\|D_n - D\| \to 0$ *as* $n \to \infty$ and $\delta_n \to 0$ *as* $n \to \infty$, we have

$$\lim_{n\to\infty} \left(\frac{T(F_0 + \delta_n D_n) - T(F_0)}{\delta_n} - T'_{F_0}(D)\right) = 0 \tag{4.23}$$

It is clear that T'_{F_0} is a linear statistical functional and $\delta_n T'_{F_0}(D) = T'_{F_0}(\delta_n D)$. Now we have the following

$$
\begin{aligned}
&\lim_{n\to\infty}\frac{T(F_0+\delta_n D_n)-T(F_0)}{\delta_n}-T'_{F_0}(D)=\lim_{n\to\infty}\frac{T(F_0+\delta_n D_n)-T(F_0)-T'_{F_0}(\delta_n D)}{\delta_n}\\
&=\lim_{n\to\infty}\frac{\int g(t,F_0+\delta_n D_n)d(F_0+\delta_n D_n)-\int g(t,F_0)dF_0-\int(g(t,F_0)+\tilde{h}(t))d(\delta_n D)}{\delta_n}\\
&=\lim_{n\to\infty}\int\frac{g(t,F_0+\delta_n D_n)-g(t,F_0)}{\delta_n}dF_0\\
&+\lim_{n\to\infty}\int\left(g(t,F_0+\delta_n D_n)-g(t,F_0)\right)d(D_n-D)\\
&+\lim_{n\to\infty}\int g(t,F_0)d(D_n-D)+\lim_{n\to\infty}\int\left(g(t,F_0+\delta_n D_n)-g(t,F_0)\right)dD-\int\tilde{h}(t)dD\\
&=(1)+(2)+(3)+(4)-(5)
\end{aligned}
\tag{4.24}
$$

Similar to the proof of Lemma 3.2.2, we can prove that (1)=(5) and (2)=(3)=(4)=0.

Then we prove that

$$\lim_{n\to\infty}\left(\frac{T(F_0+\delta_n D_n)-T(F_0)}{\delta_n}-T'_{F_0}(D)\right)=0 \tag{4.25}$$

which implies that T is Hadamard differentiable at F_0.

Since $F(0)=0$ and F is bounded, we do not need further assumptions. By the **Proposition 4.3.3** of Fernholz (1983), the remaining term is

$$\left|T(F)-T(F_0)-T'_{F_0}(F-F_0)\right|=o\left(\|F(t)-F_0(t)\|\right) \tag{4.26}$$

In particular, if $F(t)=\hat{F}_n(t)$, as is well known that $\left\|\hat{F}_n(t)-F_0(t)\right\|=O_p(\frac{1}{\sqrt{n}})$, the remaining term is $o_p(\frac{1}{\sqrt{n}})$. ∎

Above is a short discussion of Hadamard differentiability of the non-linear statistical functional $\int g(t, F(t))dF(t)$. Next, we investigate the EL with the linear statistical functional constraint $T(F_0) + T'_{F_0}(F - F_0) = \theta$.

Let $X_1, \ldots, X_n$ be n i.i.d. random variables with distribution function F. The empirical likelihood (EL) of these n observations is

$$EL(F) = \prod_{i=1}^{n} \Delta F(X_i) \tag{4.27}$$

Lemma 4.2.4 *Let $X_1, \ldots, X_n$ be n i.i.d. random variables with distribution function F_0 and the constraint below is feasible*

$$T(F_0) + T'_{F_0}(F - F_0) = \int (g(t, F_0(t)) + \tilde{h}(t))dF(t) - \int \tilde{h}(t)dF_0(t) = \theta \tag{4.28}$$

Then the maximum of EL (4.27) under the constraint (4.28) is obtained when

$$w_i = \frac{1}{n}\frac{1}{1 + \lambda(g(X_i, F_0(X_i)) + \tilde{h}(X_i) - \tilde{\theta} - \theta)} = \Delta\hat{F}_n(X_i)\frac{1}{1 + \lambda Z_i}, \;\; i = 1, \ldots, n \tag{4.29}$$

where $\hat{F}_n(t)$ is the empirical distribution function and $Z_i = g(X_i, F_0(X_i)) + \tilde{h}(X_i) - \tilde{\theta} - \theta$, $h = \frac{\partial g}{\partial F}$, $\tilde{h}(x) = \int_x^{+\infty} h(t, F_0(t))dF_0(t)$,$\tilde{\theta} = \int \tilde{h}(t)dF_0(t)$.
λ is the solution of the following equation

$$l(\lambda) = \tilde{\theta} + \theta \tag{4.30}$$

where

$$l(\lambda) = \sum_{i=1}^{n} \Delta\hat{F}_n(X_i)\frac{g(X_i, F_0(X_i)) + \tilde{h}(X_i)}{1 + \lambda(g(X_i, F_0(X_i)) + \tilde{h}(X_i) - \tilde{\theta} - \theta)} \tag{4.31}$$

Proof Denote $w_i = \Delta F(X_i)$, $EL(F) = \prod_{i=1}^n \Delta F(X_i) = \prod_{i=1}^n w_i$. Consider

$$\log EL(F) = \sum_{i=1}^n \log w_i \tag{4.32}$$

subject to the following constraint

$$w_i \geq 0,\ \sum_{i=1}^n w_i = 1,\ \sum_{i=1}^n \left(g(X_i, F_0(X_i)) + \tilde{h}(X_i)\right) w_i = \tilde{\theta} + \theta \tag{4.33}$$

In order to apply the Lagrange multiplier method, we form the target function as follows.

$$G = \sum_{i=1}^n \log w_i + \gamma \left(1 - \sum_{i=1}^n w_i\right) + n\lambda \left(\theta + \tilde{\theta} - \sum_{i=1}^n \left(g(X_i, F_0(X_i)) + \tilde{h}(X_i)\right) w_i\right) \tag{4.34}$$

Taking the derivative with respect to $w_i,\ i = 1, \ldots, n$, and equaling them to 0 yields

$$\frac{\partial G}{\partial w_i} = \frac{1}{w_i} - \gamma - n\lambda \left(g(X_i, F_0(X_i)) + \tilde{h}(X_i)\right) = 0 \tag{4.35}$$

so we have

$$\gamma = \frac{1}{w_i} - n\lambda \left(g(X_i, F_0(X_i)) + \tilde{h}(X_i)\right) \tag{4.36}$$

Multiplying w_i on both sides and taking the summation through 1 to n gives us

$$\gamma = \sum_{i=1}^n w_i \gamma = n - n\lambda \sum_{i=1}^n \left(g(X_i, F_0(X_i)) + \tilde{h}(X_i)\right) w_i = n - n\lambda \left(\tilde{\theta} + \theta\right) \tag{4.37}$$

Plug γ in (4.35), we have

$$w_i = \frac{1}{n + n\lambda\left(g(X_i, F_0(X_i)) + \tilde{h}(X_i) - \tilde{\theta} - \theta\right)}$$
$$= \frac{1}{n}\frac{1}{1 + \lambda\left(g(X_i, F_0(X_i)) + \tilde{h}(X_i) - \tilde{\theta} - \theta\right)} \tag{4.38}$$
$$= \Delta\hat{F}_n(X_i)\frac{1}{1+\lambda Z_i}$$

where $\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} I[X_i \le t]$ and $Z_i = g(X_i, F_0(X_i)) + \tilde{h}(X_i) - \tilde{\theta} - \theta$.

Plug the w_i in the constraint (4.33), we have an equation for λ as follows.

$$l(\lambda) = \sum_{i=1}^{n} \Delta\hat{F}_n(X_i)\frac{g(X_i, F_0(X_i)) + \tilde{h}(X_i)}{1 + \lambda\left(g(X_i, F_0(X_i)) + \tilde{h}(X_i) - \tilde{\theta} - \theta\right)} = \tilde{\theta} + \theta \tag{4.39}$$

A discussion of the feasibility can be found in Lemma 3.2.3. It applies to this lemma similarly. ∎

In the next lemma, we prove that the limiting distribution of $n\lambda^2$ is a $\chi^2_{(1)}$ times a constant.

Lemma 4.2.5 *Suppose $g(t, F(t))$ is left continuous with respect to t and twice differentiable with respect to F and satisfies*

$$0 < \int \left(g(t, F_0(t)) + \tilde{h}(t)\right)^2 dF_0(t) - \left(\int \left(g(t, F_0(t)) + \tilde{h}(t)\right) dF_0(t)\right)^2 < \infty \tag{4.40}$$

where $\tilde{h}(t) = \int_t^{+\infty} h(s, F_0(s))dF_0(s)$ and $h = \frac{\partial g}{\partial F}$.

Then the solution λ of (4.30) with $\theta = \theta_0$ satisfies

$$\lambda = \frac{\bar{Z}}{S_Z^2} + o_p\left(\frac{1}{\sqrt{n}}\right) \tag{4.41}$$

and

$$n\lambda^2 \xrightarrow{D} \chi^2_{(1)} \left(\int \left(g(t, F_0(t)) + \tilde{h}(t) \right)^2 dF_0(t) - \left(\int \left(g(t, F_0(t)) + \tilde{h}(t) \right) dF_0(t) \right)^2 \right)^{-1} \tag{4.42}$$

where $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$, $S_Z^2 = \frac{1}{n}\sum_{i=1}^{n} Z_i^2$, $Z_i = g(X_i, F_0(X_i)) + \tilde{h}(X_i) - \tilde{\theta} - \theta_0$, $\tilde{\theta} = \int \tilde{h}(t)dF_0(t)$ *and* $\theta_0 = \int g(t, F_0(t))dF_0(t)$.

Proof See Appendix. ∎

In the following theorem, we prove that

$$-2\log ELR(\theta_0) \xrightarrow{D} \chi^2_{(1)} \text{ as } n \to \infty \tag{4.43}$$

under the null hypothesis

$$H_0 : T(F_0) + T'_{F_0}(F - F_0) = \theta_0 \tag{4.44}$$

where $ELR(\theta_0)$ is the maximum of ELR function defined later in (4.46) subject to the linear constraint $T(F_0) + T'_{F_0}(F - F_0) = \theta_0$.

Theorem 4.2.6 *Let* $X_1, \ldots, X_n$ *be n i.i.d. random variables with distribution function* F_0. *Suppose* $g(t, F(t))$ *is left continuous with respect to* t *and twice differentiable with respect to* F *and*

$$0 < \int \left(g(t, F_0(t)) + \tilde{h}(t) \right)^2 dF_0(t) - \left(\int \left(g(t, F_0(t)) + \tilde{h}(t) \right) dF_0(t) \right)^2 < \infty \tag{4.45}$$

ELR is defined by

$$ELR(\theta) = \frac{sup\left\{ EL(F) | F \ll \hat{F}_n \text{ and } F \text{ satisfy } (4.28) \right\}}{EL(\hat{F}_n)} \tag{4.46}$$

where $\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I[X_i \le x]$, *then we have*

$$-2\log ELR(\theta_0) \xrightarrow{D} \chi^2_{(1)} \quad as \quad n \to \infty \tag{4.47}$$

Proof Consider

$$\begin{aligned} -2\log ELR(\theta_0) &= 2n\log\frac{1}{n} - 2\sum_{i=1}^{n}\log w_i = 2\sum_{i=1}^{n}\left(\log\frac{1}{n} - \log\left(\frac{1}{n}\frac{1}{1+\lambda Z_i}\right)\right) \\ &= 2\sum_{i=1}^{n}\log\left(1+\lambda Z_i\right) \end{aligned} \tag{4.48}$$

In Appendix, we have $\lambda = O_p(\frac{1}{\sqrt{n}})$ and $\max_i |Z_i| = o_p(\sqrt{n})$.

Then we have $\max_{1\le i\le n}|\lambda Z_i| = o_p(1)$, so we may expand $\log(1+\lambda Z_i)$ as follows.

$$\log\left(1+\lambda Z_i\right) = \lambda Z_i - \frac{1}{2}\lambda^2 Z_i^2 + O_p(\lambda^3)Z_i^3 \tag{4.49}$$

Therefore, we have

$$-2\log ELR(\theta_0) = 2\lambda\sum_{i=1}^{n} Z_i - \lambda^2\sum_{i=1}^{n} Z_i^2 + O_p(\lambda^3)\sum_{i=1}^{n} Z_i^3 \tag{4.50}$$

we also have

$$\begin{aligned} O_p\left(\lambda^3\right)\sum_{i=1}^{n} Z_i^3 &\le O_p\left(\lambda^3\right) \times n \times \max_i |Z_i| \times \frac{1}{n}\sum_{i=1}^{n} Z_i^2 \\ &= O_p\left(n^{-\frac{3}{2}}\right) \times n \times o_p(\sqrt{n}) \times O_p(1) = o_p(1) \end{aligned} \tag{4.51}$$

Denote $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$ and $S_Z^2 = \frac{1}{n}\sum_{i=1}^{n} Z_i^2$; in the previous lemma, we have

$$\lambda = \frac{\bar{Z}}{S_Z^2} + o_p\left(\frac{1}{\sqrt{n}}\right) \tag{4.52}$$

Plug λ in (4.50). In Appendix, we have $n\bar{Z} = O_p(\sqrt{n})$, then we have

$$\begin{aligned}
-2\log ELR(\theta_0) &= 2n\lambda\bar{Z} - n\lambda^2 S_Z^2 + o_p(1) \\
&= 2n\bar{Z}\left(\frac{\bar{Z}}{S_Z^2} + o_p(\frac{1}{\sqrt{n}})\right) - nS_Z^2\left(\frac{\bar{Z}}{S_Z^2} + o_p(\frac{1}{\sqrt{n}})\right)^2 + o_p(1) \\
&= \frac{2n\bar{Z}^2}{S_Z^2} + o_p(1) - nS_Z^2\left(\frac{\bar{Z}^2}{S_Z^4} + 2\frac{\bar{Z}}{S_Z^2}o_p(\frac{1}{\sqrt{n}}) + o_p(\frac{1}{\sqrt{n}})\right) + o_p(1) \\
&= \frac{n\bar{Z}^2}{S_Z^2} + o_p(1)
\end{aligned} \tag{4.53}$$

In Appendix, we also know that

$$\frac{n\bar{Z}^2}{S_Z^2} \xrightarrow{D} \chi^2_{(1)}, as\ n \to \infty \tag{4.54}$$

By the Slutsky theorem, we have

$$-2\log ELR(\theta_0) \xrightarrow{D} \chi^2_{(1)}, as\ n \to \infty \tag{4.55}$$

∎

Theorem 4.2.7 *Suppose the non-linear statistical functional $T(F) = \int g(t, F(t))dF(t)$ is Hadamard differentiable at $F_0(t)$. The sufficient conditions for Hadamard differentiable are listed in Theorem 4.2.3. Empirical likelihood ratio (ELR) function is defined as follows.*

$$ELR = \frac{EL(F)}{EL(\hat{F}_n)} \tag{4.56}$$

where $EL(\cdot)$ is defined in (2.3) and $\hat{F}_n$ is the empirical distribution function.
If the following null hypothesis is true

$$H_0 : \int g(t, F(t))dF(t) = \theta_0 \tag{4.57}$$

where $\theta_0 = \int g(t, F_0(t))dF_0(t)$,

then we have

$$-2\log ELR(\theta_0) \xrightarrow{D} \chi^2_{(1)} \tag{4.58}$$

where $ELR(\theta_0)$ *is the maximum of the ELR function (4.56) subject to the non-linear statistical functional constraint* $\int g(t, F(t))dF(t) = \theta_0$.

Proof This theorem is a straightforward result of Theorem 4.2.1, Theorem 4.2.3 and Theorem 4.2.6. ∎

Chapter 5 Empirical Likelihood Ratio in Terms of Cumulative Hazard for Two Sample Problems

5.1 Introduction

Hu and Barton (2009) prove that $-2\log ELR$ converges to a χ_1^2 when the following null hypothesis is true.

$$H_0: \int\int H(t,s)d\Lambda_1(t)d\Lambda_2(s) = \theta_0 \tag{5.1}$$

In this chapter, we prove the theorem in a new way using the Hadamard derivative, which would substantially simplify the proof and calculation. We also prove a new theorem involving the hazard-type null hypothesis of the following generalized form.

$$H_0: \int\int H\left(t,s,\Lambda_1(t),\Lambda_2(s)\right)d\Lambda_1(t)d\Lambda_2(s) = \theta_0 \tag{5.2}$$

We believe that the new generalized form has value in its own right. The theorem can be applied to the two-sample hypothesis like

$$H_0: P(X > Y) = p_0 \tag{5.3}$$

and the hypothesis testing of two Gini indexes equal.

To be more clear,

$$H_0: P(X > Y) = p_0 \tag{5.4}$$

is equivalent to

$$\int_0^\infty \int_0^\infty I[x > y]dF_1(x)dF_2(y) = p_0 \tag{5.5}$$

where F_1 and F_2 are distribution functions of X and Y respectively.

If the true distributions are continuous, (5.5) can be transformed to

$$\int_0^\infty \int_0^\infty I[x > y] e^{-\Lambda_1(x)} e^{-\Lambda_2(y)} d\Lambda_1(x) d\Lambda_2(y) = p_0 \tag{5.6}$$

where Λ_1 and Λ_2 are cumulative hazard functions of X and Y respectively. This is the form of our new generalization.

For the hypothesis testing of the two Gini indexes equal,

$$H_0 : G_1 = G_2 \tag{5.7}$$

is equivalent to

$$\frac{1}{\mu_1} \int_0^\infty x(1 - F_1(x)) dF_1(x) = \frac{1}{\mu_2} \int_0^\infty y(1 - F_2(y)) dF_2(y) \tag{5.8}$$

If the true distributions are continuous, (5.8) can be transformed to

$$\int_0^\infty \int_0^\infty xy e^{-\Lambda_1(x)} e^{-\Lambda_2(y)} (e^{-\Lambda_1(x)} - e^{-\Lambda_2(y)}) d\Lambda_1(x) d\Lambda_2(y) = 0 \tag{5.9}$$

which is the form of our new generalization.

5.2 A New Proof for Two Sample Problems

First of all, we propose a theorem similar to Theorem 3.2.1.

Theorem 5.2.1 *Suppose we have two statistical functional constraints in terms of cumulative hazard functions as follows*

$$T_1(\Lambda_1, \Lambda_2) = \theta_0, \quad T_2(\Lambda_1, \Lambda_2) = \theta_0 \tag{5.10}$$

which satisfy $\theta_0 = T_1(\Lambda_{10}, \Lambda_{20})$ *and* $\theta_0 = T_2(\Lambda_{10}, \Lambda_{20})$.

The ELR function is defined as follows.

$$ELR = \frac{EL(\Lambda_1, \Lambda_2)}{EL(\hat{\Lambda}_1, \hat{\Lambda}_2)} \tag{5.11}$$

where $EL(\cdot,\cdot)$ *is defined later in (5.33) and* $\hat{\Lambda}_1$, $\hat{\Lambda}_2$ *are Nelson-Aalen estimators. If for any statistics* $\hat{\Lambda}_{1n}(t)$, $\hat{\Lambda}_{2m}(s)$*(not necessary Nelson-Aalen estimators), s.t.* $\left\|\hat{\Lambda}_{1n}(t) - \Lambda_{10}(t)\right\| = O_p\left(\frac{1}{\sqrt{n}}\right)$, $\left\|\hat{\Lambda}_{2m}(s) - \Lambda_{20}(s)\right\| = O_p\left(\frac{1}{\sqrt{m}}\right)$ *and* $\frac{n}{n+m} \to \alpha$, *as* $\min(n, m) \to \infty$, *we have*

$$\left|T_1(\hat{\Lambda}_{1n}, \hat{\Lambda}_{2m}) - T_2(\hat{\Lambda}_{1n}, \hat{\Lambda}_{2m})\right| = o_p\left(\sqrt{\frac{nm}{n+m}}\right) \tag{5.12}$$

then we have

$$-2\log ELR_1(\theta_0) + 2\log ELR_2(\theta_0) \xrightarrow{P} 0, \text{ as } \min(n, m) \to \infty \tag{5.13}$$

where $ELR_1(\theta_0)$ *and* $ELR_2(\theta_0)$ *are the maximums of the ELR function (5.33) subject to the corresponding constraint* $T_1(\Lambda_1, \Lambda_2) = \theta_0$ *and* $T_2(\Lambda_1, \Lambda_2) = \theta_0$ *respectively.*

Proof Similar to the proof of Theorem 3.2.1. ∎

To begin with, we investigate the Hadamard differentiability of the statistical functional $T(\Lambda_1, \Lambda_2) = \int H(t, s)d\Lambda_1(t)d\Lambda_2(s)$ in the following theorem.

Theorem 5.2.2 *Let* $T : \mathbb{D}[0, \tau] \times \mathbb{D}[0, \tau] \longrightarrow \mathbb{R}$ *be a statistical functional defined as*

$$T(\Lambda_1(t), \Lambda_2(s)) = \int\int H(t, s)d\Lambda_1(t)d\Lambda_2(s) \tag{5.14}$$

where $\mathbb{D}[0,\tau]$ *is the set of all real valued cadlag functions equipped with sup norm.*

$$\|(f,g)\| = \max\left\{\sup_{t\in[0,\tau]}|f(t)|, \sup_{s\in[0,\tau]}|g(s)|\right\}, \qquad f,g \in \mathbb{D}[0,\tau] \tag{5.15}$$

Under some regularity conditions,

Assumption (A) $H(t,s)$ *is left continuous with respect to* t *and* s.

Assumption (B)

$$\begin{aligned} &\Lambda_1(0) = 0 \;\; and \;\; \Lambda_1(\tau) \le M_1 \; for \; some \; M_1 \; \in \mathbb{R} \\ &\Lambda_2(0) = 0 \;\; and \;\; \Lambda_2(\tau) \le M_2 \; for \; some \; M_2 \; \in \mathbb{R} \end{aligned} \tag{5.16}$$

T *is Hadamard differentiable at* $(\Lambda_{10}(t), \Lambda_{20}(s))$ *with derivative*

$$\begin{aligned} &T'_{\Lambda_{10},\Lambda_{20}}(\Lambda_1(t)-\Lambda_{10}(t), \Lambda_2(s)-\Lambda_{20}(s)) \\ &= \int H_1(t)d(\Lambda_1(t)-\Lambda_{10}(t)) + \int H_2(s)d(\Lambda_2(s)-\Lambda_{20}(s)) \end{aligned} \tag{5.17}$$

where $H_1(t) = \int H(t,s)d\Lambda_{20}(s)$ *and* $H_2(s) = \int H(t,s)d\Lambda_{10}(t)$. *If* $\Lambda_1(t) = \hat{\Lambda}_1(t)$, $\Lambda_2(s) = \hat{\Lambda}_2(s)$, *where* $\hat{\Lambda}_1$ *and* $\hat{\Lambda}_2$ *are Nelson-Aalen estimators, the remaining term*

$$\left|T(\Lambda_1(t),\Lambda_2(s)) - T(\Lambda_{10}(t),\Lambda_{20}(s)) - T'_{\Lambda_{10},\Lambda_{20}}(\Lambda_1(t)-\Lambda_{10}(t), \Lambda_2(s)-\Lambda_{20}(s))\right| \tag{5.18}$$

is $o_p\left(\sqrt{\frac{n+m}{nm}}\right)$, *where* n *and* m *are sample sizes of* $\hat{\Lambda}_1$ *and* $\hat{\Lambda}_2$ *respectively.*

Proof First we prove that (5.15) is a norm. The definition of the norm is as follows:

Definition Given a vector space $\mathbb{V}$ over a field $\mathbb{F}$, a norm on $\mathbb{V}$ is a function $p : \mathbb{V} \to \mathbb{R}$ with the following properties: For all $a \in \mathbb{F}$ and all $u, v \in \mathbb{V}$,

- $p(av) = |a|v$

- $p(u+v) \le p(u) + p(v)$

- If $p(v) = 0$, the v is the zero vector.

It is easy to verify that the $\|\cdot\|$ defined in the Lemma 5.2.2 is a norm. It is obvious for the first and third rule. We would prove the second rule here.
Suppose $(f_1, g_1), (f_2, g_2) \in D[0,\tau] \times D[0,\tau]$, we have

$$\begin{aligned}
&\|(f_1, g_1) + (f_2, g_2)\| = \|(f_1 + f_2, g_1 + g_2)\| = \max\left\{\sup\left\{|f_1 + f_2|\right\}, \sup\left\{|g_1 + g_2|\right\}\right\} \\
&\le \max\left\{\sup|f_1| + \sup|f_2|, \sup|g_1| + \sup|g_2|\right\} \\
&\le \max\left\{\sup|f_1|, \sup|g_1|\right\} + \max\left\{\sup|f_2|, \sup|g_2|\right\} \\
&= \|(f_1, g_1)\| + \|(f_2, g_2)\|
\end{aligned} \tag{5.19}$$

Next we are going to prove that T is Hadamard differentiable at $(\Lambda_{10}(t), \Lambda_{20}(s))$ with Hadamard derivative

$$T'_{\Lambda_{10},\Lambda_{20}} = \int H_1(t) d\left(\Lambda_1(t) - \Lambda_{10}(t)\right) + \int H_2(s) d\left(\Lambda_2(s) - \Lambda_{20}(s)\right) \tag{5.20}$$

where $H_1(t) = \int H(t,s) d\Lambda_{20}(s)$ and $H_2(s) = \int H(t,s) d\Lambda_{10}(t)$.

In order to prove this, we need to prove that $\boldsymbol{\Lambda_0} = (\Lambda_{10}(t), \Lambda_{20}(s))$, $\boldsymbol{D} = (D_1(t), D_2(s))$, $\boldsymbol{D_n} = (D_{1n}(t), D_{2n}(s))$ and $\boldsymbol{D}, \boldsymbol{D_n} \in \mathbb{D}[0,\tau] \times \mathbb{D}[0,\tau]$, s.t. $\|\boldsymbol{D_n} - \boldsymbol{D}\| \to 0$ as $n \to \infty$ and $\delta_n \to 0$ as $n \to \infty$, we have

$$\lim_{n\to\infty} \left(\frac{T(\boldsymbol{\Lambda_0} + \delta_n \boldsymbol{D_n}) - T(\boldsymbol{\Lambda_0})}{\delta_n} - T'_{\Lambda_{10},\Lambda_{20}}(\boldsymbol{D}) \right) = 0 \tag{5.21}$$

To save typing, unless otherwise specified, Λ_{10}, Λ_{20}, D_{1n}, D_1, D_{2n}, D_2 represent $\Lambda_{10}(t)$, $\Lambda_{20}(s)$, $D_{1n}(t)$, $D_1(t)$, $D_{2n}(s)$, $D_2(s)$ respectively.
Note that $T'_{\Lambda_{10},\Lambda_{20}}$ is a linear statistical functional so that $\delta_n T'_{\Lambda_{10},\Lambda_{20}}(\boldsymbol{D}) = T'_{\Lambda_{10},\Lambda_{20}}(\delta_n \boldsymbol{D})$.

Then we have

$$
\begin{aligned}
&\lim_{n\to\infty}\left(\frac{T(\boldsymbol{\Lambda_0}+\delta_n\boldsymbol{D_n})-T(\boldsymbol{\Lambda_0})}{\delta_n}-T^{'}_{\Lambda_{10},\Lambda_{20}}(\boldsymbol{D})\right)\\
&=\lim_{n\to\infty}\frac{T(\boldsymbol{\Lambda_0}+\delta_n\boldsymbol{D_n})-T(\boldsymbol{\Lambda_0})-T^{'}_{\Lambda_{10},\Lambda_{20}}(\delta_n\boldsymbol{D})}{\delta_n}\\
&=\lim_{n\to\infty}(\frac{\int\int H(t,s)d\left(\Lambda_{10}+\delta_nD_{1n}\right)d\left(\Lambda_{20}+\delta_nD_{2n}\right)-\int\int H(t,s)d\Lambda_{10}(t)d\Lambda_{20}(s)}{\delta_n}\\
&-\frac{\int\int H(t,s)d\Lambda_{20}d\left(\delta_nD_1\right)-\int\int H(t,s)d\Lambda_{10}d\left(\delta_nD_2\right)}{\delta_n})\\
&=\lim_{n\to\infty}\int\int H(t,s)d\Lambda_{20}d\left(D_{1n}-D_1\right)+\lim_{n\to\infty}\int\int H(t,s)d\Lambda_{10}d\left(D_{2n}-D_2\right)\\
&+\lim_{n\to\infty}\delta_n\int\int H(t,s)dD_{1n}dD_{2n}\\
&=\lim_{n\to\infty}\int H_1(t)d\left(D_{1n}(t)-D_1(t)\right)+\lim_{n\to\infty}\int H_2(s)d\left(D_{2n}(s)-D_2(s)\right)\\
&+\lim_{n\to\infty}\delta_n\int\int H(t,s)dD_{1n}dD_{2n}\\
&=\lim_{n\to\infty}\left(H_1(t)\left(D_{1n}(t)-D_1(t)\right)|_0^\tau-\int\left(D_{1n}(t)-D_1(t)\right)dH_1(t)\right)\\
&+\lim_{n\to\infty}\left(H_2(s)\left(D_{2n}(s)-D_2(s)\right)|_0^\tau-\int\left(D_{2n}(s)-D_2(s)\right)dH_2(s)\right)\\
&+\lim_{n\to\infty}\delta_n\int\int H(t,s)dD_{1n}dD_{2n}
\end{aligned}
\tag{5.22}
$$

We also have that $D_{1n}(t)$, $D_{2n}(s)$ converge pointwisely to $D_1(t)$, $D_2(s)$ as $n\to\infty$ since the following.

$$
\begin{aligned}
&|D_{1n}(t)-D_1(t)|\le\sup_{t\in[0,\tau]}|D_{1n}(t)-D_1(t)|\\
&\le\max\left\{\sup_{t\in[0,\tau]}|D_{1n}(t)-D_1(t)|,\ \sup_{sin[0,\tau]}|D_{2n}(s)-D_2(s)|\right\}\\
&=\|\boldsymbol{D_n}-\boldsymbol{D}\|\to 0\quad as\ \ n\to\infty
\end{aligned}
\tag{5.23}
$$

Similarly, $|D_{2n}(s)-D_2(s)|\to 0\quad as\ \ n\to\infty$

In particular, we have

$$
\begin{aligned}
&|D_{1n}(0) - D_1(0)| \to 0 \;\; as \;\; n \to \infty, \quad |D_{2n}(0) - D_2(0)| \to 0, \;\; as \;\; n \to \infty \\
&|D_{1n}(\tau) - D_1(\tau)| \to 0 \;\; as \;\; n \to \infty, \quad |D_{2n}(\tau) - D_2(\tau)| \to 0, \;\; as \;\; n \to \infty
\end{aligned}
\tag{5.24}
$$

Then we have

$$
\begin{aligned}
&\lim_{n\to\infty} H_1(t)\left(D_{1n}(t) - D_1(t)\right)|_0^\tau = 0 \\
&\lim_{n\to\infty} H_2(s)\left(D_{2n}(s) - D_2(s)\right)|_0^\tau = 0
\end{aligned}
\tag{5.25}
$$

We have already learned that D_{1n} and D_{2n} are uniformly bounded,

so $\int\int H(t,s)dD_{1n}dD_{2n}$ is uniformly bounded as well. Therefore we have

$$
\lim_{n\to\infty} \delta_n \int\int H(t,s)dD_{1n}dD_{2n} = 0 \tag{5.26}
$$

By the integration by parts, we have

$$
\begin{aligned}
&\int H_1(t)d\left(D_{1n}(t) - D_1(t)\right) = H_1(t)\left(D_{1n}(t) - D_1(t)\right)|_0^\tau - \int \left(D_{1n}(t) - D_1(t)\right)dH_1(t) \\
&\int H_2(s)d\left(D_{2n}(s) - D_2(s)\right) = H_2(s)\left(D_{2n}(s) - D_2(s)\right)|_0^\tau - \int \left(D_{2n}(s) - D_2(s)\right)dH_2(s)
\end{aligned}
\tag{5.27}
$$

Since D_{1n} and D_1 are uniformly bounded, $|D_{1n} - D_1|$ is uniformly bounded as well. By the Dominated Convergence Theorem, we have

$$
\begin{aligned}
\lim_{n\to\infty}\left|\int \left(D_{1n}(t) - D_1(t)\right)dH_1(t)\right| &\le \lim_{n\to\infty}\int \left|\left(D_{1n}(t) - D_1(t)\right)\right| d\left|H_1(t)\right| \\
&= \int \lim_{n\to\infty}\left|\left(D_{1n}(t) - D_1(t)\right)\right| d\left|H_1(t)\right| = 0
\end{aligned}
\tag{5.28}
$$

Similarly, we have

$$\begin{aligned}\lim_{n\to\infty}\left|\int (D_{2n}(s)-D_2(s))\,dH_2(s)\right| &\le \lim_{n\to\infty}\int |(D_{2n}(s)-D_2(s))|\,d\,|H_2(s)| \\ &= \int \lim_{n\to\infty}|(D_{2n}(s)-D_2(s))|\,d\,|H_2(s)| = 0\end{aligned} \tag{5.29}$$

Now we prove that (5.22)=0.

With further assumptions that

$$\begin{aligned}&\Lambda_1(0)=0 \ \ and \ \ \Lambda_1(\tau)\le M_1 \ for\ some\ M_1 \ \in \mathbb{R} \\ &\Lambda_2(0)=0 \ \ and \ \ \Lambda_2(\tau)\le M_2 \ for\ some\ M_2 \ \in \mathbb{R}\end{aligned} \tag{5.30}$$

and $\Lambda_1(t)=\hat{\Lambda}_1(t)$, $\Lambda_2(s)=\hat{\Lambda}_2(s)$, where $\hat{\Lambda}_1(t)$ and $\hat{\Lambda}_2(s)$ are Nelson-Aalen estimators. By Fernholz (1983) Proposition 4.3.4, the remaining term

$$\left|T(\Lambda_1(t),\Lambda_2(s)) - T(\Lambda_{10}(t),\Lambda_{20}(s)) - T'_{\Lambda_{10},\Lambda_{20}}(\Lambda_1(t)-\Lambda_{10}(t),\Lambda_2(s)-\Lambda_{20}(s))\right| \tag{5.31}$$

is $o_p\left(\sqrt{\frac{n+m}{nm}}\right)$. ∎

In Theorem 5.2.2, we learned that the statistical functional $T(\Lambda_1,\Lambda_2)=\int H(t,s)d\Lambda_1(t)d\Lambda_2(s)$ is Hadamard differentiable with a linear Hadamard derivative. This is of central importance, because now we can obtain an explicit maximum of cumulative hazard function using Lagrange multiplier method. We state this in the following lemma.

To start with, we introduce the empirical likelihood of two samples.

Suppose $(X_1,\dots,X_n)$, $(Y_1,\dots,Y_m)$ are n and m i.i.d. random variables with distribution functions F_1 and F_2 denoting lifetimes, respectively. $(C_1,\dots,C_n)$, $(D_1,\dots,D_m)$ are n and m i.i.d. random variables with distribution functions G_1 and G_2, independent of X and Y and denoting censoring times, respectively. X and Y are independent. C and D are independent. And only the censored observations are available to us.

$$
\begin{aligned}
T_i &= \min(X_i, C_i), \quad \delta_{X_i} = I[X_i \le C_i], \quad i = 1, \ldots, n \\
U_j &= \min(Y_j, D_j), \quad \delta_{Y_j} = I[Y_j \le D_j], \quad j = 1, \ldots, m
\end{aligned}
\tag{5.32}
$$

The empirical likelihood (EL) of $\Lambda_1(t)$ and $\Lambda_2(s)$ which is dominated by the Nelson-Aalen estimator is

$$
\begin{aligned}
EL = &\prod_{i=1}^{n} \left\{ w_i^{\delta_{X_i}} \left[\prod_{k:T_k<T_i} (1-w_k) \right]^{\delta_{X_i}} \left[\prod_{k:T_k\le T_i} (1-w_k) \right]^{1-\delta_{X_i}} \right\} \\
&\times \prod_{j=1}^{m} \left\{ v_j^{\delta_{Y_j}} \left[\prod_{k:U_k<U_j} (1-v_k) \right]^{\delta_{Y_j}} \left[\prod_{k:U_k\le U_j} (1-v_k) \right]^{1-\delta_{Y_j}} \right\}
\end{aligned}
\tag{5.33}
$$

where $w_i = \Delta\Lambda_1(T_i)$ and $v_j = \Delta\Lambda_2(U_j)$.

The Poisson extension of the empirical likelihood is

$$
AL = \prod_{i=1}^{n} w_i^{\delta_{X_i}} \exp\left\{ -\sum_{k=1}^{i} w_k \right\} \prod_{j=1}^{m} v_j^{\delta_{Y_j}} \exp\left\{ -\sum_{k=1}^{j} v_k \right\}
\tag{5.34}
$$

Lemma 5.2.3 *If the constraint below is feasible*

$$
\int H_1(t) d\Lambda_1(t) + \int H_2(s) d\Lambda_2(s) = 2\theta
\tag{5.35}
$$

then the maximum of AL (5.34) under the above constraint is obtained when

$$
w_i = \Delta\hat{\Lambda}_1(T_i) \frac{1}{1+\lambda Z_{1i}}, \quad v_j = \Delta\hat{\Lambda}_2(U_j) \frac{1}{1+\lambda Z_{2j}}, \quad i = 1, \ldots, n-1, \quad j = 1, \ldots, m-1
\tag{5.36}
$$

where $Z_{1i} = \frac{\delta_{X_i} H_1(T_i)}{n-i+1}$ *and* $Z_{2j} = \frac{\delta_{Y_j} H_2(U_j)}{m-j+1}$, *i=1,. . . , n, j=1,. . . ,m*

$$H_1(t) = \int H(t,s)d\Lambda_{20}(s), \quad H_2(s) = \int H(t,s)d\Lambda_{10}(t) \tag{5.37}$$

λ is the solution of the following equation

$$l(\lambda) = 2\theta \tag{5.38}$$

where

$$l(\lambda) = \sum_{i=1}^{n-1} \frac{Z_{1i}}{1+\lambda Z_{1i}} + Z_{1n} + \sum_{j=1}^{m-1} \frac{Z_{2j}}{1+\lambda Z_{2j}} + Z_{2m} \tag{5.39}$$

Proof Notice that $w_n = \delta_{X_n}$ and $v_m = \delta_{Y_m}$, we can rewrite the constraint in the discrete form as below.

$$\sum_{i=1}^{n-1} \delta_{X_i} H_1(T_i) w_i + \delta_{X_n} H_1(T_n) + \sum_{j=1}^{m-1} \delta_{Y_j} H_2(U_j) v_j + \delta_{Y_m} H_2(U_m) = 2\theta \tag{5.40}$$

The $\log AL$ is

$$\begin{aligned} \log AL &= \sum_{i=1}^{n} \left(\delta_{X_i} \log w_i - \sum_{l=1}^{i} w_l \right) + \sum_{j=1}^{m} \left(\delta_{Y_j} \log v_j - \sum_{k=1}^{j} v_k \right) \\ &= \sum_{i=1}^{n} \delta_{X_i} \log w_i - \sum_{i=1}^{n} (n-i+1)\, w_i + \sum_{j=1}^{m} \delta_{Y_j} \log v_j - \sum_{j=1}^{m} (m-j+1)\, v_j \end{aligned} \tag{5.41}$$

To use the Lagrange multiplier method, we form the following target function.

$$\begin{aligned} G = &\sum_{i=1}^{n} \delta_{X_i} \log w_i - \sum_{i=1}^{n} (n-i+1)\, w_i + \sum_{j=1}^{m} \delta_{Y_j} \log v_j - \sum_{j=1}^{m} (m-j+1)\, v_j \\ &- \lambda \left(\delta_{X_i} H_1(T_i) w_i + \delta_{X_n} H_1(T_n) + \sum_{j=1}^{m-1} \delta_{Y_j} H_2(U_j) v_j + \delta_{Y_m} H_2(U_m) - 2\theta \right) \end{aligned} \tag{5.42}$$

Taking the derivative with respect to $w_i, i = 1, \ldots, n-1$ and $v_j, j = 1, \ldots, m-1$ and equaling them to 0 yields

$$
\begin{aligned}
\frac{\partial G}{\partial w_i} &= \frac{\delta_{X_i}}{w_i} - (n-i+1) - \lambda \delta_{X_i} H_1(T_i) = 0, \quad i = 1, \dots, n-1 \\
\frac{\partial G}{\partial v_j} &= \frac{\delta_{Y_j}}{v_j} - (m-j+1) - \lambda \delta_{Y_j} H_2(U_j) = 0, \quad j = 1, \dots, m-1
\end{aligned}
\tag{5.43}
$$

so the jumps of Λ_1 and Λ_2 at times T_i and U_j are

$$
\begin{aligned}
w_i = \Delta\Lambda_1(T_i) &= \frac{\delta_{X_i}}{n-i+1} \frac{1}{1+\lambda \frac{\delta_{X_i} H_1(T_i)}{n-i+1}}, \quad i = 1, \dots, n-1 \\
v_j = \Delta\Lambda_2(U_j) &= \frac{\delta_{Y_j}}{m-j+1} \frac{1}{1+\lambda \frac{\delta_{Y_j} H_2(U_j)}{m-j+1}}, \; j = 1, \dots, m-1
\end{aligned}
\tag{5.44}
$$

Let

$$
\begin{aligned}
Z_{1i} &= \frac{\delta_{X_i} H_1(T_i)}{n-i+1}, \quad i = 1, \dots, n \\
Z_{2j} &= \frac{\delta_{Y_j} H_2(U_j)}{m-j+1}, \quad j = 1, \dots, m
\end{aligned}
\tag{5.45}
$$

We have

$$
\begin{aligned}
w_i &= \Delta\hat{\Lambda}_1(T_i) \frac{1}{1+\lambda Z_{1i}}, \quad i = 1, \dots, n-1, \quad w_n = \delta_{X_n} \\
v_j &= \Delta\hat{\Lambda}_2(U_j) \frac{1}{1+\lambda Z_{2j}}, \quad j = 1, \dots, m-1, \quad v_m = \delta_{Y_m}
\end{aligned}
\tag{5.46}
$$

where $\Delta\hat{\Lambda}_1$ and $\Delta\hat{\Lambda}_2$ are Nelson-Aalen estimators.

Plug the $w_i, \; i = 1, \dots, n-1, \;\; v_j, \; j = 1, \dots, m-1, \; w_n, \;\; v_m$ into the discrete form of the constraint (5.40), we have the following equation

$$
\begin{aligned}
&\sum_{i=1}^{n-1} \delta_{X_i} H_1(T_i) \frac{\delta_{X_i}}{n-i+1} \frac{1}{1+\lambda Z_{1i}} + \delta_{X_n} H_1(T_i) \\
&+ \sum_{j=1}^{m-1} \delta_{Y_j} H_2(U_j) \frac{\delta_{Y_j}}{m-j+1} \frac{1}{1+\lambda Z_{2j}} + \delta_{Y_m} H_2(U_m) = 2\theta
\end{aligned}
\tag{5.47}
$$

which can be simplified as below

$$\sum_{i=1}^{n-1} \frac{Z_{1i}}{1+\lambda Z_{1i}} + Z_{1n} + \sum_{j=1}^{m-1} \frac{Z_{2j}}{1+\lambda Z_{2j}} + Z_{2m} = 2\theta \tag{5.48}$$

∎

In the following lemma, we prove that the limiting distribution of $\frac{n+m}{nm}\lambda^2$ is a $\chi^2_{(1)}$ times a constant.

Lemma 5.2.4 *Suppose $H(t,s)$ is left continuous with respect to t and s and we have*

$$\begin{aligned} &\int \frac{H_1^2(t)d\Lambda_{10}(t)}{(1-F_1(t))(1-G_1(t))} < \infty \\ &\int \frac{H_2^2(s)d\Lambda_{20}(s)}{(1-F_2(s))(1-G_2(s))} < \infty \end{aligned} \tag{5.49}$$

We assume that

$$\frac{n}{n+m} \to \alpha, \ \ as \ \min(n,m) \to \infty \tag{5.50}$$

The solution λ of (5.39) with $\theta = \theta_0$ satisfies

$$\frac{n+m}{nm}\lambda^2 \xrightarrow{D} \chi_1^2 \cdot \sigma^{-2} \quad as \quad \min(n,m) \to \infty \tag{5.51}$$

where

$$\sigma^2 = (1-\alpha)\int \frac{H_1^2(t)d\Lambda_{10}(t)}{(1-F_1(t))\,(1-G_1(t))} + \alpha \int \frac{H_2^2(s)d\Lambda_{20}(s)}{(1-F_2(s))\,(1-G_2(s))} \tag{5.52}$$

Proof See Appendix. ∎

Lemma 5.2.5 *Let $(T_1, \delta_{X_1}), \ldots, (T_n, \delta_{X_n})$ and $(U_1, \delta_{Y_1}), \ldots, (U_m, \delta_{Y_m})$ be n and m pairs of random variables as defined before. Suppose $H(t,s)$ is left continuous with* *respect to t and s and we have*

$$\int \frac{H_1^2(t)d\Lambda_{10}(t)}{(1-F_1(t))(1-G_1(t))} < \infty \ , \quad \int \frac{H_2^2(s)d\Lambda_{20}(s)}{(1-F_2(s))(1-G_2(s))} < \infty \tag{5.53}$$

ALR defined by

$$ALR(\theta) = \frac{\sup\left\{AL(\Lambda_1, \Lambda_2)|\Lambda_1 \ll \hat{\Lambda}_1, \Lambda_2 \ll \hat{\Lambda}_2 \ and \ \Lambda_1 \ , \Lambda_2 \ \ satisfy \ \ (5.2)\right\}}{AL(\hat{\Lambda}_1, \hat{\Lambda}_2)} \tag{5.54}$$

Then

$$-2logALR(\theta_0) \xrightarrow{D} \chi^2_{(1)} \quad as \ \ n \to \infty \tag{5.55}$$

Proof Since the Poisson extension of the empirical likelihood is

$$AL = \prod_{i=1}^{n} w_i^{\delta_{X_i}} \exp\left\{-\sum_{k=1}^{i} w_k\right\} \prod_{j=1}^{m} v_j^{\delta_{Y_j}} \exp\left\{-\sum_{k=1}^{j} v_k\right\} \tag{5.56}$$

where $w_i = \Delta\Lambda_1(T_i)$, $v_j = \Delta\Lambda_2(U_j)$.

Take log on both sides, we have

$$\begin{aligned} \log AL &= \sum_{i=1}^{n} \delta_{X_i} \log w_i - \sum_{i=1}^{n}\sum_{k=1}^{i} w_k + \sum_{j=1}^{m} \delta_{Y_j} \log v_j - \sum_{j=1}^{m}\sum_{k=1}^{j} v_k \\ &= \sum_{i=1}^{n} \delta_{X_i} \log w_i - \sum_{i=1}^{n} (n-i+1)\, w_i + \sum_{j=1}^{m} \delta_{Y_j} \log v_j - \sum_{j=1}^{m} (m-j+1)\, v_j \end{aligned} \tag{5.57}$$

Define

$$\begin{aligned} Z_{1i} &= \frac{\delta_{X_i} H_1(T_i)}{n-i+1}, \quad i = 1, \dots, n \\ Z_{2j} &= \frac{\delta_{Y_j} H_2(U_j)}{m-j+1}, \quad j = 1, \dots, m \end{aligned} \tag{5.58}$$

Consider

$$\begin{aligned}
&-2\log ALR(\theta_0) = 2\left[\sum_{i=1}^{n}\delta_{X_i}\log\left(\Delta\hat{\Lambda}_1(T_i)\right) - \sum_{i=1}^{n}(n-i+1)\,\Delta\hat{\Lambda}_1(T_i)\right] \\
&-2\left[\sum_{i=1}^{n-1}\delta_{X_i}\log\left(\Delta\hat{\Lambda}_1(T_i)\frac{1}{1+\lambda Z_{1i}}\right) - \sum_{i=1}^{n-1}(n-i+1)\,\Delta\hat{\Lambda}_1(T_i)\frac{1}{1+\lambda Z_{1i}}\right] \\
&-2\left[\delta_{X_n}\log\left(\Delta\hat{\Lambda}_1(T_n)\right) - \Delta\hat{\Lambda}_1(T_n)\right] \\
&+2\left[\sum_{j=1}^{m}\delta_{Y_j}\log\left(\Delta\hat{\Lambda}_2(U_j)\right) - \sum_{j=1}^{m}(m-j+1)\,\Delta\hat{\Lambda}_2(U_j)\right] \\
&-2\left[\sum_{j=1}^{m-1}\delta_{Y_j}\log\left(\Delta\hat{\Lambda}_2(U_j)\frac{1}{1+\lambda Z_{2j}}\right) - \sum_{j=1}^{m-1}(m-j+1)\,\Delta\hat{\Lambda}_2(U_j)\frac{1}{1+\lambda Z_{2j}}\right] \\
&-2\left[\delta_{U_m}\log\left(\Delta\hat{\Lambda}_2(U_m)\right) - \Delta\hat{\Lambda}_2(U_m)\right] \\
&= 2\sum_{i=1}^{n-1}\delta_{X_i}\log(1+\lambda Z_{1i}) - 2\sum_{i=1}^{n-1}(n-i+1)\,\Delta\hat{\Lambda}_1(T_i)\frac{\lambda Z_{1i}}{1+\lambda Z_{1i}} \\
&+2\sum_{j=1}^{m-1}\delta_{Y_j}\log(1+\lambda Z_{2j}) - 2\sum_{j=1}^{m-1}(m-j+1)\,\Delta\hat{\Lambda}_2(U_j)\frac{\lambda Z_{2j}}{1+\lambda Z_{2j}} \\
&= 2\sum_{i=1}^{n-1}\delta_{X_i}\log(1+\lambda Z_{1i}) - 2\sum_{i=1}^{n-1}\frac{\delta_{X_i}\lambda Z_{1i}}{1+\lambda Z_{1i}} \\
&+2\sum_{j=1}^{m-1}\delta_{Y_j}\log(1+\lambda Z_{2j}) - 2\sum_{j=1}^{m-1}\frac{\delta_{Y_j}\lambda Z_{2j}}{1+\lambda Z_{2j}} \\
&= 2\sum_{i=1}^{n-1}\delta_{X_i}\log(1+\lambda Z_{1i}) - 2\sum_{i=1}^{n-1}\delta_{X_i}\lambda Z_{1i} + 2\sum_{i=1}^{n-1}\frac{\delta_{X_i}\lambda^2 Z_{1i}^2}{1+\lambda Z_{1i}} \\
&+2\sum_{j=1}^{m-1}\delta_{Y_j}\log(1+\lambda Z_{2j}) - 2\sum_{j=1}^{m-1}\delta_{Y_j}\lambda Z_{2j} + 2\sum_{j=1}^{m-1}\frac{\delta_{U_j}\lambda^2 Z_{2j}^2}{1+\lambda Z_{2j}}
\end{aligned} \tag{5.59}$$

In Appendix, notice that

$$\begin{aligned}
\max_{1\le i\le n}|\lambda Z_{1i}| &= |\lambda|\max_{1\le i\le n}|Z_{1i}| = O_p\left(\frac{1}{\sqrt{n}}\right)o_p\left(\sqrt{n}\right) = o_p(1) \\
\max_{1\le j\le m}|\lambda Z_{2j}| &= |\lambda|\max_{1\le j\le m}|Z_{2j}| = O_p\left(\frac{1}{\sqrt{n}}\right)o_p\left(\sqrt{n}\right) = o_p(1)
\end{aligned} \tag{5.60}$$

We may expand $\log(1+\lambda Z_{1i})$ and $\log(1+\lambda Z_{2j})$ as

$$\begin{aligned}\log(1+\lambda Z_{1i}) &= \lambda Z_{1i} - \frac{1}{2}\lambda^2 Z_{1i}^2 + O_p\left(\lambda^3\right) Z_{1i}^3 \\ \log(1+\lambda Z_{2j}) &= \lambda Z_{2j} - \frac{1}{2}\lambda^2 Z_{2j}^2 + O_p\left(\lambda^3\right) Z_{2j}^3\end{aligned} \tag{5.61}$$

Substituting this into (5.59) gives us

$$\begin{aligned}-2\log ALR(\theta_0) &= \sum_{i=1}^{n-1}\delta_{X_i}\lambda^2 Z_{1i}^2 + 2O_p\left(\lambda^3\right)\sum_{i=1}^{n-1} Z_{1i}^3 - 2\sum_{i=1}^{n-1}\frac{\delta_{X_i}\lambda^3 Z_{1i}^3}{1+\lambda Z_{1i}} \\ &+ \sum_{j=1}^{m-1}\delta_{Y_j}\lambda^2 Z_{2j}^2 + 2O_p\left(\lambda^3\right)\sum_{j=1}^{m-1} Z_{2j}^3 - 2\sum_{j=1}^{m-1}\frac{\delta_{Y_j}\lambda^3 Z_{2j}^3}{1+\lambda Z_{2j}} \\ &= \frac{n+m}{nm}\lambda^2\left(\frac{m}{n+m}n\sum_{i=1}^{n-1} Z_{1i}^2 + \frac{n}{n+m}m\sum_{j=1}^{m-1} Z_{2j}^2\right) \\ &+ 2O_p\left(\lambda^3\right)\sum_{i=1}^{n-1} Z_{1i}^3 - 2\sum_{i=1}^{n-1}\frac{\delta_{X_i}\lambda^3 Z_{1i}^3}{1+\lambda Z_{1i}} \\ &+ 2O_p\left(\lambda^3\right)\sum_{j=1}^{m-1} Z_{2j}^3 - 2\sum_{j=1}^{m-1}\frac{\delta_{Y_j}\lambda^3 Z_{2j}^3}{1+\lambda Z_{2j}}\end{aligned} \tag{5.62}$$

Similar to the proof in Lemma 3.2.5, we have

$$\begin{aligned}&\left|O_p\left(\lambda^3\right)\sum_{i=1}^{n-1} Z_{1i}^3\right| = o_p(1), \quad \left|O_p\left(\lambda^3\right)\sum_{j=1}^{m-1} Z_{2j}^3\right| = o_p(1) \\ &\sum_{i=1}^{n-1}\frac{\delta_{X_i}\lambda^3 Z_{1i}^3}{1+\lambda Z_{1i}} = o_p(1), \quad \sum_{j=1}^{m-1}\frac{\delta_{Y_j}\lambda^3 Z_{2j}^3}{1+\lambda Z_{2j}} = o_p(1)\end{aligned} \tag{5.63}$$

We assume that $\frac{n}{n+m} \to \alpha,\ as\ \min(n,m) \to \infty$. In Appendix, we have

$$\begin{aligned}&\frac{n+m}{nm}\lambda^2 \xrightarrow{D} \chi^2_{(1)}\sigma^{-2} \\ &\frac{m}{n+m}n\sum_{i=1}^{n-1} Z_{1i}^2 + \frac{n}{n+m}m\sum_{j=1}^{m-1} Z_{2j}^2 \xrightarrow{P} \sigma^2\end{aligned} \tag{5.64}$$

where

$$\sigma^2 = (1-\alpha)\int \frac{H_1^2(t)d\Lambda_{10}(t)}{(1-F_1(t))\,(1-G_1(t))} + \alpha \int \frac{H_2^2(s)d\Lambda_{20}(s)}{(1-F_2(s))\,(1-G_2(s))} \tag{5.65}$$

By the Slutsky theorem, we have

$$-2\log ALR(\theta_0) \xrightarrow{D} \chi^2_{(1)},\ as\ n \to \infty \tag{5.66}$$

$\blacksquare$

Theorem 5.2.6 *Suppose all conditions of Lemma 5.2.5 hold and ELR is defined by*

$$ELR(\theta) = \frac{EL(\Lambda_1^*, \Lambda_2^*)}{EL(\hat{\Lambda}_1, \hat{\Lambda}_2)} \tag{5.67}$$

where $\hat{\Lambda}_1$ *and* $\hat{\Lambda}_2$ *are Nelson-Aalen estimators and* Λ_1^* *and* Λ_1^* *are given by the jumps defined in Lemma 5.2.3*

Then we have,

$$-2logELR(\theta_0) \xrightarrow{D} \chi^2_{(1)} \qquad as\ n \to \infty \tag{5.68}$$

Proof Remember that the empirical likelihood for two samples is

$$\begin{aligned} EL = \prod_{i=1}^{n} &\left\{ w_i^{\delta_{X_i}} \left[\prod_{k:T_k<T_i} (1-w_k) \right]^{\delta_{X_i}} \left[\prod_{k:T_k \le T_i} (1-w_k) \right]^{1-\delta_{X_i}} \right\} \\ \times \prod_{j=1}^{m} &\left\{ v_j^{\delta_{Y_j}} \left[\prod_{k:U_k<U_j} (1-v_k) \right]^{\delta_{Y_j}} \left[\prod_{k:U_k \le U_j} (1-v_k) \right]^{1-\delta_{Y_j}} \right\} \end{aligned} \tag{5.69}$$

where $w_i = \Delta\Lambda_1(T_i)$ and $v_j = \Delta\Lambda_2(U_j)$.

$$\log EL = \sum_{i=1}^{n} \delta_{X_i} \log w_i + \sum_{i=1}^{n} (n-i+1-\delta_{X_i}) \log(1-w_i) + \sum_{j=1}^{m} \delta_{Y_j} \log v_j + \sum_{j=1}^{m} \left(m-j+1-\delta_{Y_j}\right) \log(1-v_j) \tag{5.70}$$

Similar to the proof of Theorem 3.2.6, we have

$$-2\log ELR(\theta_0) = 2\sum_{i=1}^{n-1} \delta_{X_i} \log(1+\lambda Z_{1i}) - 2\sum_{i=1}^{n-1} \frac{\delta_{X_i}\lambda Z_{1i}}{1+\lambda Z_{1i}} + 2\sum_{i=1}^{n-1} \frac{\delta_{X_i}\eta_{1i}^2}{(n-i+1-\delta_{X_i})} + 2\sum_{j=1}^{m-1} \delta_{Y_j} \log(1+\lambda Z_{2j}) - 2\sum_{j=1}^{m-1} \frac{\delta_{Y_j}\lambda Z_{2j}}{1+\lambda Z_{2j}} + 2\sum_{j=1}^{m-1} \frac{\delta_{Y_j}\eta_{2j}^2}{\left(m-j+1-\delta_{Y_j}\right)} \tag{5.71}$$

where $|\eta_{1i}| \leq |\frac{\lambda Z_{1i}}{1+\lambda Z_{1i}}|$ and $|\eta_{2j}| \leq |\frac{\lambda Z_{2j}}{1+\lambda Z_{2j}}|$. Remember that

$$-2\log ALR(\theta_0) = 2\sum_{i=1}^{n-1} \delta_{X_i} \log(1+\lambda Z_{1i}) - 2\sum_{i=1}^{n-1} \frac{\delta_{X_i}\lambda Z_{1i}}{1+\lambda Z_{1i}} + 2\sum_{j=1}^{m-1} \delta_{Y_j} \log(1+\lambda Z_{2j}) - 2\sum_{j=1}^{m-1} \frac{\delta_{Y_j}\lambda Z_{2j}}{1+\lambda Z_{2j}} \tag{5.72}$$

We have

$$-2\log ELR(\theta_0) + 2\log ALR(\theta_0) = 2\sum_{i=1}^{n-1} \frac{\delta_{X_i}\eta_{1i}^2}{(n-i+1-\delta_{X_i})} + 2\sum_{j=1}^{m-1} \frac{\delta_{Y_j}\eta_{2j}^2}{\left(m-j+1-\delta_{Y_j}\right)} \tag{5.73}$$

Similar to the proof of Theorem 3.2.6, we have

$$2\sum_{i=1}^{n-1} \frac{\delta_{X_i}\eta_{1i}^2}{(n-i+1-\delta_{X_i})} = o_p(1), \qquad 2\sum_{j=1}^{m-1} \frac{\delta_{Y_j}\eta_{2j}^2}{\left(m-j+1-\delta_{Y_j}\right)} = o_p(1) \tag{5.74}$$

Therefore

$$-2\log ELR(\theta_0) + 2\log ALR(\theta_0) \xrightarrow{P} 0 \tag{5.75}$$

By Slutsky theorem,

$$-2\log ELR(\theta_0) \xrightarrow{D} \chi^2_{(1)} \; as \; n \; \to \infty \tag{5.76}$$

∎

Theorem 5.2.7 *Suppose the non-linear statistical functional* $T(\Lambda_1, \Lambda_2) = \int H(t,s)d\Lambda_1(t)d\Lambda_2(s)$ *is Hadamard differentiable at* $(\Lambda_{10}(t), \Lambda_{20}(s))$. *The sufficient conditions for Hadamard differentiability are listed in Theorem 5.2.2. The empirical likelihood ratio (ELR) function is defined as follows.*

$$ELR = \frac{EL(\Lambda_1, \Lambda_2)}{EL(\hat{\Lambda}_1, \hat{\Lambda}_2)} \tag{5.77}$$

where $EL(\cdot,\cdot)$ *is defined in (5.33) and* $\hat{\Lambda}_1$, $\hat{\Lambda}_2$ *are the Nelson-Aalen estimators. If the following null hypothesis is true*

$$H_0 : \int H(t,s)d\Lambda_1(t)d\Lambda_2(s) = \theta_0 \tag{5.78}$$

where $\theta_0 = \int H(t,s)d\Lambda_{10}(t)d\Lambda_{20}(s)$,
then we have

$$-2\log ELR(\theta_0) \xrightarrow{D} \chi^2_{(1)} \tag{5.79}$$

where $ELR(\theta_0)$ *is the maximum of the ELR function (5.77) subject to the non-linear statistical functional constraint* $\int H(t,s)d\Lambda_1(t)d\Lambda_2(s) = \theta_0$.

Proof This theorem is a straightforward result of Theorem 5.2.1, Theorem 5.2.2 and Theorem 5.2.6. ∎

5.3 A New Generalization of Two Sample Problems

In this section, we prove the theorem under a new generalized hazard-type form of null hypothesis as follows.

$$H_0 : \int\int H\left(t, s, \Lambda_1(t), \Lambda_2(s)\right) d\Lambda_1(t) d\Lambda_2(s) = \theta_0 \tag{5.80}$$

Still we begin with the Hadamard differentiability of the statistical functional constraint in the following theorem.

Theorem 5.3.1 *Let $T : \mathbb{D}[0,\tau] \times \mathbb{D}[0,\tau] \to \mathbb{R}$ be a statistical functional defined as*

$$T(\Lambda_1(t), \Lambda_2(s)) = \int\int H(t, s, \Lambda_1(t), \Lambda_2(s)) d\Lambda_1(t) d\Lambda_2(s) \tag{5.81}$$

where $\mathbb{D}[0,\tau]$ is the set of real valued cadlag functions equipped with the sup norm.

$$\|(f,g)\| = \max\left\{\sup_{t\in[0,\tau]} |f(t)|, \sup_{s\in[0,\tau]} |g(s)|\right\}, \qquad f, g \in \mathbb{D}[0,\tau] \tag{5.82}$$

Under some regularity conditions,

Assumption (A) *$H\left(t, s, \Lambda_1(t), \Lambda_2(s)\right)$ is left continuous with respect to t and s and twice differentiable with respect to Λ_1 and Λ_2.*

Assumption (B) *$H\left(t, s, \Lambda_1(t), \Lambda_2(s)\right)$, H_1 and H_2 are bounded by integrable functions $A(t,s)$, $B(t,s)$ and $C(t,s)$ respectively, for any $t, s \in [0,\tau]$ and $\Lambda_1(t), \Lambda_2(s) \in \mathbb{D}[0,\tau]$*

Assumption (C)

$$\begin{aligned} \Lambda_1(0) = 0 \ \ \textit{and} \ \ \Lambda_1(\tau) \le M_1 \ \textit{for some} \ M_1 \in \mathbb{R} \\ \Lambda_2(0) = 0 \ \ \textit{and} \ \ \Lambda_2(\tau) \le M_2 \ \textit{for some} \ M_2 \in \mathbb{R} \end{aligned} \tag{5.83}$$

T is Hadamard differentiable at $(\Lambda_{10}(t), \Lambda_{20}(s))$ *with derivative*

$$\begin{aligned} &T'_{\Lambda_{10},\Lambda_{20}}\left(\Lambda_1(t)-\Lambda_{10}(t), \Lambda_2(s)-\Lambda_{20}(s)\right) \\ &= \int H_1^*(t) d\left(\Lambda_1(t)-\Lambda_{10}(t)\right) + \int H_2^*(s) d\left(\Lambda_2(s)-\Lambda_{20}(s)\right) \end{aligned} \tag{5.84}$$

where

$$\begin{aligned} H_1^*(t) &= \int \Big(H\left(t,s,\Lambda_{10}(t),\Lambda_{20}(s)\right) + \tilde{H}_1\left(t,s,\Lambda_{20}(s)\right)\Big) d\Lambda_{20}(s) \\ H_2^*(s) &= \int \Big(H(t,s,\Lambda_{10}(t),\Lambda_{20}(s)) + \tilde{H}_2(t,s,\Lambda_{10}(t)\Big) d\Lambda_{10}(t) \\ \tilde{H}_1(t,s,\Lambda_{20}(s)) &= \int_t^{+\infty} H_1(x,s,\Lambda_{10}(x),\Lambda_{20}(s)) d\Lambda_{10}(x) \\ \tilde{H}_2(t,s,\Lambda_{10}(t)) &= \int_s^{+\infty} H_2(t,y,\Lambda_{10}(t),\Lambda_{20}(y)) d\Lambda_{20}(y) \\ H_1 &= \frac{\partial H}{\partial \Lambda_1}, \qquad H_2 = \frac{\partial H}{\partial \Lambda_2} \end{aligned} \tag{5.85}$$

If $\Lambda_1(t) = \hat{\Lambda}_1(t)$ *and* $\Lambda_2(s) = \hat{\Lambda}_2(s)$, *where* $\hat{\Lambda}_1(t)$ *and* $\hat{\Lambda}_2(s)$ *are Nelson-Aalen estimators, the remaining term is*

$$\left|T(\Lambda_1,\Lambda_2) - T(\Lambda_{10},\Lambda_{20}) - T'_{(\Lambda_{10},\Lambda_{20})}\left(\Lambda_1-\Lambda_{10},\Lambda_2-\Lambda_{20}\right)\right| = o_p\left(\sqrt{\frac{n+m}{nm}}\right) \tag{5.86}$$

where n *and* m *are sample sizes of* $\hat{\Lambda}_1(t)$ *and* $\hat{\Lambda}_2(s)$ *respectively.*

Proof In order to prove that T is Hadamard differentiable, let $\boldsymbol{\Lambda_0} = (\Lambda_{10}(t), \Lambda_{20}(s))$, $\boldsymbol{D} = (D_1(t), D_2(s))$, $\boldsymbol{D_n} = (D_{1n}(t), D_{2n}(s))$ and $\boldsymbol{D}, \boldsymbol{D_n} \in \mathbb{D}[0,\tau] \times \mathbb{D}[0,\tau]$, s.t. $\|\boldsymbol{D_n} - \boldsymbol{D}\| \to 0$ as $n \to \infty$ and $\delta_n \to 0$ as $n \to \infty$, we have

$$\lim_{n\to\infty} \frac{T(\boldsymbol{\Lambda_0} + \delta_n \boldsymbol{D_n}) - T(\boldsymbol{\Lambda_0})}{\delta_n} - T'_{\Lambda_{10},\Lambda_{20}}(\boldsymbol{D}) = 0 \tag{5.87}$$

To save typing, unless otherwise specified, Λ_{10}, Λ_{20}, D_{1n}, D_{2n}, D_1, D_2, H, $\tilde{H}_1$, $\tilde{H}_2$ represent $\Lambda_{10}(t)$, $\Lambda_{20}(s)$, $D_{1n}(t)$, $D_{2n}(s)$, $D_1(t)$, $D_2(s)$, $H(t,s,\Lambda_{10}(t),\Lambda_{20}(s))$,$\tilde{H}_1(t,s,\Lambda_{20}(s))$,

$\tilde{H}_2(t, s, \Lambda_{10}(t))$ respectively.

Note that $T'_{\Lambda_{10},\Lambda_{20}}$ is a linear statistical functional. Then we have

$$
\begin{aligned}
&\lim_{n\to\infty}\left(\frac{T(\boldsymbol{\Lambda_0}+\delta_n\boldsymbol{D_n})-T(\boldsymbol{\Lambda_0})}{\delta_n}-T'_{\Lambda_{10},\Lambda_{20}}(\boldsymbol{D})\right)\\
&=\lim_{n\to\infty}(\frac{\int\int H\left(t,s,\Lambda_{10}+\delta_nD_{1n},\Lambda_{20}+\delta_nD_{2n}\right)d\left(\Lambda_{10}+\delta_nD_{1n}\right)d\left(\Lambda_{20}+\delta_nD_{2n}\right)}{\delta_n}\\
&-\frac{\int\int Hd\Lambda_{10}d\Lambda_{20}}{\delta_n})\\
&-\int\int\left(H+\tilde{H}_1\right)dD_1d\Lambda_{20}-\int\int\left(H+\tilde{H}_2\right)d\Lambda_{10}dD_2\\
&=\lim_{n\to\infty}\int\int\frac{H\left(t,s,\Lambda_{10}+\delta_nD_{1n},\Lambda_{20}+\delta_nD_{2n}\right)-H\left(t,s,\Lambda_{10},\Lambda_{20}+\delta_nD_{2n}\right)}{\delta_n}d\Lambda_{10}d\Lambda_{20}\\
&+\lim_{n\to\infty}\int\int\frac{H\left(t,s,\Lambda_{10},\Lambda_{20}+\delta_nD_{2n}\right)-H\left(t,s,\Lambda_{10},\Lambda_{20}\right)}{\delta_n}d\Lambda_{10}d\Lambda_{20}\\
&+\lim_{n\to\infty}\int\int\left(H\left(t,s,\Lambda_{10}+\delta_nD_{1n},\Lambda_{20}+\delta_nD_{2n}\right)-H\left(t,s,\Lambda_{10},\Lambda_{20}\right)\right)d\left(D_{1n}-D_1\right)d\Lambda_{20}\\
&+\lim_{n\to\infty}\int\int\left(H\left(t,s,\Lambda_{10}+\delta_nD_{1n},\Lambda_{20}+\delta_nD_{2n}\right)-H\left(t,s,\Lambda_{10},\Lambda_{20}\right)\right)d\left(D_{2n}-D_2\right)d\Lambda_{10}\\
&+\lim_{n\to\infty}\int\int\left(H\left(t,s,\Lambda_{10}+\delta_nD_{1n},\Lambda_{20}+\delta_nD_{2n}\right)-H\left(t,s,\Lambda_{10},\Lambda_{20}\right)\right)d\left(D_1\right)d\Lambda_{20}\\
&+\lim_{n\to\infty}\int\int\left(H\left(t,s,\Lambda_{10}+\delta_nD_{1n},\Lambda_{20}+\delta_nD_{2n}\right)-H\left(t,s,\Lambda_{10},\Lambda_{20}\right)\right)d\left(D_2\right)d\Lambda_{10}\\
&+\lim_{n\to\infty}\int\int H\left(t,s,\Lambda_{10},\Lambda_{20}\right)d\left(D_{1n}-D_1\right)d\Lambda_{20}\\
&+\lim_{n\to\infty}\int\int H\left(t,s,\Lambda_{10},\Lambda_{20}\right)d\left(D_{2n}-D_2\right)d\Lambda_{10}\\
&-\int\int H_1\left(t,s,\Lambda_{10},\Lambda_{20}\right)D_1d\Lambda_{10}d\Lambda_{20}-\int\int H_2\left(t,s,\Lambda_{10},\Lambda_{20}\right)D_2d\Lambda_{10}d\Lambda_{20}\\
&=0
\end{aligned}
\tag{5.88}
$$

Similar arguments of proof of Lemma 3.2.2 can prove the above calculation.

With further assumptions as follows

$$
\begin{aligned}
&\Lambda_1(0)=0\ \ and\ \ \Lambda_1(\tau)\le M_1\ for\ some\ M_1\ \in\mathbb{R}\\
&\Lambda_2(0)=0\ \ and\ \ \Lambda_2(\tau)\le M_2\ for\ some\ M_2\ \in\mathbb{R}
\end{aligned}
\tag{5.89}
$$

and $\Lambda_1(t) = \hat{\Lambda}_1(t)$, $\Lambda_2(s) = \hat{\Lambda}_2(s)$, where $\hat{\Lambda}_1(t)$ and $\hat{\Lambda}_2(s)$ are Nelson-Aalen estimators, by Fernholz (1983) Proposition 4.3.4, the remaining term

$$\left| T(\Lambda_1(t), \Lambda_2(s)) - T(\Lambda_{10}(t), \Lambda_{20}(s)) - T'_{\Lambda_{10},\Lambda_{20}}(\Lambda_1(t) - \Lambda_{10}(t), \Lambda_2(s) - \Lambda_{20}(s)) \right| \tag{5.90}$$

is $o_p\left(\sqrt{\frac{n+m}{nm}}\right)$, where n and m are sample sizes of $\hat{\Lambda}_1(t)$ and $\hat{\Lambda}_2(s)$ respectively. ∎

Suppose $(T_1, \delta_{X_1}), \ldots, (T_n, \delta_{X_n})$ and $(U_1, \delta_{Y_1}), \ldots, (U_m, \delta_{Y_m})$ are n and m i.i.d. random vectors defined in (5.32).

Lemma 5.3.2 *If the constraint below is feasible*

$$\int H_1^*(t) d\Lambda_1(t) + \int H_2^*(s) d\Lambda_2(s) - \tilde{\theta}_1 - \tilde{\theta}_2 = 2\theta \tag{5.91}$$

where $\tilde{\theta}_1 = \int\int \tilde{H}_1 d\Lambda_{10} d\Lambda_{20}$ *and* $\tilde{\theta}_2 = \int\int \tilde{H}_2 d\Lambda_{10} d\Lambda_{20}$*;* $\tilde{H}_1$*,* $\tilde{H}_2$*,* H_1^* *and* H_2^* *are defined in (5.85).*

Then the maximum of AL (5.34) under the above constraint is obtained when

$$w_i = \Delta\hat{\Lambda}_1(T_i)\frac{1}{1+\lambda Z_{1i}}, \;\; v_j = \Delta\hat{\Lambda}_2(U_j)\frac{1}{1+\lambda Z_{2j}}, \;\; i = 1, \ldots, n-1, \;\; j = 1, \ldots, m-1 \tag{5.92}$$

where $Z_{1i} = \frac{\delta_{X_i} H_1^*(T_i)}{n-i+1}$ *and* $Z_{2j} = \frac{\delta_{Y_j} H_2^*(U_j)}{m-j+1}$*,* $i=1,\ldots, n \quad j=1,\ldots,m$

λ *is the solution of the following equation*

$$l(\lambda) = 2\theta \tag{5.93}$$

where

$$l(\lambda) = \sum_{i=1}^{n-1} \frac{Z_{1i}}{1+\lambda Z_{1i}} + Z_{1n} + \sum_{j=1}^{m-1} \frac{Z_{2j}}{1+\lambda Z_{2j}} + Z_{2m} - \tilde{\theta}_1 - \tilde{\theta}_2 \tag{5.94}$$

Proof Notice that $w_n = \delta_{X_n}$ and $v_m = \delta_{Y_m}$. Suppose $w_i = \Delta\Lambda_1(T_i),\; i = 1, \ldots, n-1$ and $v_j = \Delta\Lambda_2(U_j),\; j = 1, \ldots, m-1$, the constraint can be written in the discrete form as follows.

$$\sum_{i=1}^{n-1} \delta_{X_i} H_1^*(T_i) w_i + \delta_{X_n} H_1^*(T_n) + \sum_{j=1}^{m-1} \delta_{Y_j} H_2^*(U_j) v_j + \delta_{Y_m} H_2^*(U_m) - \tilde{\theta}_1 - \tilde{\theta}_2 = 2\theta \quad (5.95)$$

In order to use the Lagrange multiplier method, we form the target function as below.

$$\begin{aligned} G = & \sum_{i=1}^{n} \left(\delta_{X_i} log w_i - \sum_{l=1}^{i} w_l \right) + \sum_{j=1}^{m} \left(\delta_{Y_j} log v_j - \sum_{k=1}^{j} w_k \right) \\ & - \lambda \left[\sum_{i=1}^{n-1} \delta_{X_i} H_1^*(T_i) w_i + \delta_{X_n} H_1^*(T_n) + \sum_{j=1}^{m-1} \delta_{Y_j} H_2^*(U_j) v_j + \delta_{Y_m} H_2^*(U_m) - \tilde{\theta}_1 - \tilde{\theta}_2 - 2\theta \right] \end{aligned} \quad (5.96)$$

Taking the derivative with respect to $w_i, i = 1, \ldots, n-1$ and $v_j, j = 1, \ldots, m-1$ and equaling them to 0, we have

$$\begin{aligned} \frac{\partial G}{\partial w_i} &= \frac{\delta_{X_i}}{w_i} - (n-i+1) - \lambda \delta_{X_i} H_1^*(T_i) = 0, \quad i = 1, \ldots, n-1 \\ \frac{\partial G}{\partial v_j} &= \frac{\delta_{Y_j}}{v_j} - (m-j+1) - \lambda \delta_{Y_j} H_2^*(U_j) = 0, \quad j = 1, \ldots, m-1 \end{aligned} \quad (5.97)$$

so we have

$$\begin{aligned} w_i &= \frac{\delta_{X_i}}{n-i+1} \frac{1}{1+\lambda Z_{1i}} = \Delta\hat{\Lambda}_1(T_i) \frac{1}{1+\lambda Z_{1i}}, \quad i = 1, \ldots, n-1 \\ v_j &= \frac{\delta_{Y_j}}{m-j+1} \frac{1}{1+\lambda Z_{2j}} = \Delta\hat{\Lambda}_2(U_j) \frac{1}{1+\lambda Z_{2j}}, \quad j = 1, \ldots, m-1 \end{aligned} \quad (5.98)$$

where $Z_{1i} = \frac{\delta_{X_i} H_1^*(T_i)}{n-i+1}$, $Z_{2j} = \frac{\delta_{Y_j} H_2^*(U_j)}{m-j+1}$ and $\hat{\Lambda}_1$, $\hat{\Lambda}_2$ are Nelson-Aalen estimators. Plugging the $w_i,\ i = 1, \ldots, n-1,\ w_n,\ v_j,\ j = 1, \ldots, m-1,\ v_m$ into the discrete format of the constraint we have the following equation for λ,

$$\sum_{i=1}^{n-1} \delta_{X_i} H_1^*(T_i) \frac{\delta_{X_i}}{n-i+1} \frac{1}{1+\lambda Z_{1i}} + \delta_{X_n} H_1^*(T_n) + \sum_{j=1}^{m-1} \delta_{Y_j} H_2^*(U_j) \frac{\delta_{Y_j}}{m-j+1} \frac{1}{1+\lambda Z_{2j}}$$
$$+ \delta_{Y_m} H_2^*(U_m) - \tilde{\theta}_1 - \tilde{\theta}_2 - 2\theta = 0 \tag{5.99}$$

which can be simplified as

$$\sum_{i=1}^{n-1} \frac{Z_{1i}}{1+\lambda Z_{1i}} + Z_{1n} + \sum_{j=1}^{m-1} \frac{Z_{2j}}{1+\lambda Z_{2j}} + Z_{2m} - \tilde{\theta}_1 - \tilde{\theta}_2 - 2\theta = 0 \tag{5.100}$$

$\blacksquare$

The following lemma is similar to Lemma 5.2.4.

Lemma 5.3.3 *Suppose $H(t, s, \Lambda_1(t), \Lambda_2(s))$ is continuous with respect to t and s and twice differentiable with respect to Λ_1 and Λ_2 and*

$$\int \frac{(H_1^*(t))^2 d\Lambda_{10}(t)}{(1-F_1(t))\,(1-G_1(t))} < \infty$$
$$\int \frac{(H_2^*(s))^2 d\Lambda_{20}(s)}{(1-F_2(s))\,(1-G_2(s))} < \infty \tag{5.101}$$

We assume that

$$\frac{n}{n+m} \to \alpha, \text{ as } n \to \infty \tag{5.102}$$

The solution λ of (5.94) with $\theta = \theta_0$ satisfies

$$\frac{n+m}{nm}\lambda^2 \xrightarrow{D} \chi_1^2 \cdot \sigma^{-2} \quad as \quad \min(n,m) \to \infty \tag{5.103}$$

$$\sigma^2 = (1-\alpha) \int \frac{H_1^2(t) d\Lambda_{10}(t)}{(1-F_1(t))\,(1-G_1(t))} + \alpha \int \frac{H_2^2(s) d\Lambda_{20}(s)}{(1-F_2(s))\,(1-G_2(s))} \tag{5.104}$$

Proof Similar to the proof of Lemma 5.2.4 ∎

Lemma 5.3.4 *Let* $(T_1, \delta_{X_1}), \ldots, (T_n, \delta_{X_n})$ *and* $(U_1, \delta_{Y_1}), \ldots, (U_m, \delta_{Y_m})$ *be* n *and* m *pairs of random variables as defined in (5.32). Suppose* $H(t, s, \Lambda_1(t), \Lambda_2(s))$ *is continuous with respect to* t *and* s *and twice differentiable with respect to* Λ_1 *and* Λ_2 *and we have*

$$\int \frac{(H_1^*(t))^2 d\Lambda_{10}(t)}{(1 - F_1(t))\,(1 - G_1(t))} < \infty, \qquad \int \frac{(H_2^*(s))^2 d\Lambda_{20}(s)}{(1 - F_2(s))\,(1 - G_2(s))} < \infty \tag{5.105}$$

ALR defined by

$$ALR(\theta) = \frac{sup\left\{AL(\Lambda_1, \Lambda_2)|\Lambda_1 \ll \hat{\Lambda}_1, \Lambda_2 \ll \hat{\Lambda}_2 \;\; and \;\; \Lambda_1, \Lambda_2 \;\; satisfy \;\; (5.91)\right\}}{AL(\hat{\Lambda}_1, \hat{\Lambda}_2)} \tag{5.106}$$

Then

$$-2 \log ALR(\theta_0) \xrightarrow{D} \chi^2_{(1)} \quad as \;\; n \to \infty \tag{5.107}$$

Proof Similar to the proof of Lemma 5.2.5 ∎

Theorem 5.3.5 *Suppose all conditions of Lemma 5.3.4 hold and ELR is defined by*

$$ELR(\theta) = \frac{EL(\Lambda_1^*, \Lambda_2^*)}{EL(\hat{\Lambda}_1, \hat{\Lambda}_2)} \tag{5.108}$$

where $\hat{\Lambda}_1$ *and* $\hat{\Lambda}_2$ *are Nelson-Aalen estimators and* Λ_1^* *and* Λ_1^* *are given by the jumps defined in Lemma 5.3.2*

Then we have,

$$-2 \log ELR(\theta_0) \xrightarrow{D} \chi^2_{(1)} \quad as\; n \to \infty \tag{5.109}$$

Proof Similar to the proof of Theorem 5.2.6 ∎

Theorem 5.3.6 *Suppose the non-linear statistical functional*

$T(\Lambda_1, \Lambda_2) = \int H(t, s, \Lambda_1(t), \Lambda_2(s)) d\Lambda_1(t) d\Lambda_2(s)$ *is Hadamard differentiable* *at* $(\Lambda_{10}(t), \Lambda_{20}(s))$. *The sufficient conditions for Hadamard differentiability are listed in Theorem 5.3.1. The empirical likelihood ratio (ELR) function is defined as follows.*

$$ELR = \frac{EL(\Lambda_1, \Lambda_2)}{EL(\hat{\Lambda}_1, \hat{\Lambda}_2)} \tag{5.110}$$

where $EL(\cdot, \cdot)$ *is defined in (5.33) and* $\hat{\Lambda}_1$, $\hat{\Lambda}_2$ *are the Nelson-Aalen estimators.*
If the following null hypothesis is true

$$H_0 : \int H(t, s, \Lambda_1(t), \Lambda_2(s)) d\Lambda_1(t) d\Lambda_2(s) = \theta_0 \tag{5.111}$$

where $\theta_0 = \int H(t, s, \Lambda_1(t), \Lambda_2(s)) d\Lambda_{10}(t) d\Lambda_{20}(s)$,
then we have

$$-2 \log ELR(\theta_0) \xrightarrow{D} \chi^2_{(1)} \tag{5.112}$$

where $ELR(\theta_0)$ *is the maximum of the ELR function (5.110) subject to the non-linear statistical functional constraint* $\int H(t, s, \Lambda_1(t), \Lambda_2(s)) d\Lambda_1(t) d\Lambda_2(s) = \theta_0$.

Proof This theorem is a straightforward result of Theorem 5.2.1, Theorem 5.3.1 and Theorem 5.3.5. ∎

Chapter 6 Algorithm and Simulations

6.1 Algorithm

In order to calculate the $\Lambda(t) \ll \hat{\Lambda}_{NA}$ that maximizes AL subject to the following constraint

$$\int g(t, \Lambda(t)) d\Lambda(t) = \theta_0 \tag{6.1}$$

we use the plug-in method and calculate the maximum iteratively. If we plug $\hat{\Lambda}_{NA}$ into $g\left(t, \Lambda(t)\right)$, g becomes a function that only depends on t but not on $\Lambda(t)$ any more. For this kind of g, the computation of maximum cumulative hazard functions is solved by Pan and Zhou (2002).

The algorithm we used can be summarized as follows.

Step 1 Use the Nelson-Aalen estimator($\Lambda^0(t)$) as the initial plug-in value for $\Lambda(t)$ in $g(t, \Lambda(t))$, and solve the maximization with constraint $\int g(t, \Lambda^0(t)) d\Lambda(t) = \theta_0$ to obtain $\Lambda^1(t)$.

Step 2 In n^{th} iteration, $n = 2, \ldots$, plug $\Lambda^{n-1}(t)$ into $g(t, \Lambda(t))$ of the constraint.

$$\int g(t, \Lambda^{n-1}(t)) d\Lambda(t) = \theta_0 \tag{6.2}$$

Let $\Lambda^n(t)$ denote the cumulative hazard that maximizes AL subject to the above constraint.

Step 3 Repeat Step 2 until $\| \Lambda^n - \Lambda^{n+1} \| < 10^{-12}$. For here, the $\| \cdot \|$ represents L_2 norm.

This algorithm converges in some examples, but there are cases that it does not converge, depending on the function $g(\cdot, \cdot)$.

The software used in our simulations is R version 3.1.3. The R codes of the simulations are listed in the Appendix. At the end of each simulation section, we give one R code example and track the time of the simulation. I run the R codes on my personal computer: Macbook Pro, 2.7GHz Intel Core i7, 4GB 1333 MHz DDR3.

6.2 Simulation 1

In this simulation, we compare the coverage probabilities of the confidence intervals based on EL method with other available methods: Normal Approximation, Log Transformation $y = \log(x)$, Log-log Transformation $y = \log(-\log x)$ and Arcsin Transformation $y = arcsin(\sqrt{x})$. We list the formulas of the $100\,(1-\alpha)\,\%$ confidence intervals for survival probability $S(t)$ for a fixed t as below.
Normal approximation:

$$\hat{S}(t) \pm z_{\alpha/2}\hat{S}(t)\hat{\sigma}(t) \tag{6.3}$$

Log-transformation

$$\hat{S}(t)\exp\{\pm z_{\alpha/2}\frac{\hat{\sigma}(t)}{\hat{S}(t)}\} \tag{6.4}$$

Log(-log)-transformation:

$$\hat{S}(t)^{\exp\{\pm z_{\alpha/2}\frac{\hat{\sigma}(t)}{\log\hat{S}(t)}\}} \tag{6.5}$$

Arcsin-transformation

$$\begin{aligned}
&\sin^2\left\{\max\left[0, \arcsin((\hat{S}(t))^{\frac{1}{2}} - \frac{1}{2}z_{\alpha/2}\hat{\sigma}(t)\left[\frac{\hat{S}(t)}{1-\hat{S}(t)}\right]^{\frac{1}{2}}\right]\right\}\\
&\leq S(t) \leq \\
&\sin^2\left\{\min\left[\frac{\pi}{2}, \arcsin((\hat{S}(t))^{\frac{1}{2}} - \frac{1}{2}z_{\alpha/2}\hat{\sigma}(t)\left[\frac{\hat{S}(t)}{1-\hat{S}(t)}\right]^{\frac{1}{2}}\right]\right\}
\end{aligned} \tag{6.6}$$

where $\hat{S}(t)$ is the Kaplan-Meier estimator and $\hat{\sigma}^2(t)$ is from the well-known Greenwood formula.

$$\hat{Var}(\hat{S}(t)) = \hat{S}^2(t)\hat{\sigma}^2(t) = \hat{S}^2(t) \sum_{t_i \le t} \frac{d_i}{n_i(n_i - d_i)} \tag{6.7}$$

Next, we illustrate how to generate confidence intervals based on EL method. If the null hypothesis $H_0 : \int g(t, \Lambda(t))d\Lambda(t) = \theta_0$ is true, we have,

$$-2 \log ELR(\theta_0) \xrightarrow{D} \chi_1^2, \ as \ \ n \to \infty \tag{6.8}$$

The definition of $ELR(\theta_0)$ can be found in Theorem 3.2.7.

So a $100\,(1-\alpha)\,\%$ empirical likelihood ratio confidence interval is

$$\left\{\theta \ \mid \ \theta \ \ s.t. \ \ -2\log ELR(\theta) \le \chi^2_{(1),1-\alpha}\right\} \tag{6.9}$$

where $\chi^2_{(1),1-\alpha}$ is the $(1-\alpha)\,th$ percentile of $\chi^2_{(1)}$.

Suppose we simulate one sample with sample size $n = 50$ from $F_0(t) = 1 - \exp(-t)$ and $G_0(t) = 1 - \exp(-0.35t)$, where F_0 and G_0 denote lifetimes and censoring times, respectively. By calculating $-2\log ELR(\theta)$ for various θ, we get a U-shape plot of $-2\log ELR(\theta)$ versus θ if we place θ in ascending order with small enough steps. The 95% confidence interval for θ_0 is the set of θ with value of $-2\log ELR(\theta)$ under 3.84, which is the 95th percentile of $\chi^2_{(1)}$. See Figure 6.1 for an illustration.

Now we can compare the coverage probabilities of the confidence intervals of $F(t)$ based on EL method with other four methods on various sample size. We simulate the data with various sample size $n = 50, 100, 200, 500, 1000$ from $F_0(t) = 1-\exp(-t)$ and $G_0(t) = 1 - \exp(-0.35t)$, where F_0 and G_0 denote lifetimes and censoring times, respectively. For each sample size, we generated 5000 confidence intervals.

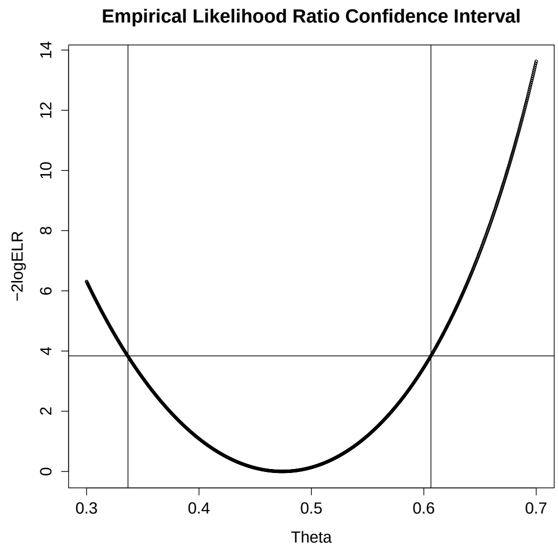

Figure 6.1: 95% Confidence Interval based on EL

n	ELR	Normal	Log	Log(-log)	Arcsin
50	0.9408	0.9392	0.9806	0.9526	0.9480
100	0.9470	0.9472	0.9894	0.9570	0.9520
200	0.9486	0.9514	0.994	0.9526	0.9516
500	0.9498	0.9516	0.9958	0.9506	0.9520
1000	0.9492	0.9512	0.9978	0.9534	0.9518

Table 6.1: Coverage Probabilities of Nominal 95% Confidence Intervals of $F(0.5)$

From Table 6.1, we see that Log transformation is not good here; EL method and arcsin transformation are similar, both of which are better than the normal approximation. However, for the real problem, we do not know which transformation is the best transformation and it seems that EL method implicitly chooses the best transformation for us.

```
### R codes of one example of Simulation 1 ###
### The comparison of the coverage probability of confidence interval
### for F(0.5) with sample size 1000 and 5000 samples ###
time1<-proc.time()
CP(1000,5000,0.5,g1,1-exp(-0.5),0.05,0.3,0.5,0.001,0.1)
```

n		ELR	Normal	Log	Log(-log)	Arcsin
50	Coverage Probability	0.95	0.95	0.97	0.97	0.95
	Average Length	0.2588	0.2813	0.4854	0.2788	0.2773
100	Coverage Probability	0.93	0.93	0.98	0.94	0.93
	Average Length	0.1892	0.1987	0.3319	0.1979	0.1973
200	Coverage Probability	0.96	0.96	1.00	0.98	0.97
	Average Length	0.1370	0.1419	0.2362	0.1416	0.1414
500	Coverage Probability	0.97	0.97	1	0.98	0.97
	Average Length	0.0872	0.0897	0.1484	0.0896	0.0896
1000	Coverage Probability	0.94	0.94	1.00	0.94	0.94
	Average Length	0.0616	0.0635	0.1052	0.0635	0.0635

Table 6.2: Coverage Probability and Average Length of Nominal 95% Confidence Intervals of $F(0.5)$

```
proc.time()-time1

### The time of simulation (in seconds). ###

5433.204
```

6.3 Simulation 2

In order to compare our method with the arcsin transformation method, we calculate the average length of the confidence intervals. Again, we simulate the data with various sample size $n = 50, 100, 200, 500, 1000$ from $F_0(t) = 1 - \exp(-t)$ and $G_0(t) = 1 - \exp(-0.35t)$, where F_0 and G_0 denote lifetimes and censoring times respectively. For each sample size, we generate 100 confidence intervals and calculate the average length of them. The results are in the Table 6.2. The first line of each cell in Table 6.2 is the coverage probability while the second line displays average length of 100 confidence intervals.

From Table 6.2, we can see that the confidence intervals based on EL method have the smallest average length among all five methods for all sample sizes. In addition, when the sample size is small ($n = 50$), the average length of confidence intervals based on EL method is significantly shorter than those of arcsin transformation confidence intervals.

```
### R codes of one example of Simulation 2 ###
### The comparison of the average length of 100 confidence intervals
### for F(0.5) with sample size 1000 ###
time1<-proc.time()
CPandAVL(1000,100,0.5,g1,1-exp(-0.5),0.05,0.3,0.5,0.001,0.1)
proc.time()-time1
### The time of simulation (in seconds). ###
66587.999
```

6.4 Simulation 3

In order to verify that the limiting distribution of $-2\log ELR(\theta_0)$ is truly $\chi^2_{(1)}$, we generate QQ plots for different sample sizes. We plot the sample quantiles of $-2\log ELR(\theta_0)$ versus the theoretical quantiles of $\chi^2_{(1)}$. If the points align along the 45 degree line, it means that the sample of $-2\log ELR(\theta_0)$ is truly $\chi^2_{(1)}$ distributed. We generate the QQ plots both when null hypothesis is true and when alternative hypothesis is true. Since there are infinite alternative hypothesis, we just choose one to illustrate the point that the limiting distribution of $-2\log ELR(\theta_0)$ is not $\chi^2_{(1)}$ when the alternative hypothesis is true.

The following list is the information of our simulation set-up.

- $F_0(t) = 1 - e^{-t}$, $G_0(t) = 1 - e^{-0.35t}$ denote lifetimes and censoring times respectively.
- $g_1 = g(t, \Lambda(t)) = I[t <= 0.5]e^{-\Lambda(t)}$.
- Approximately 26% of the data are censored.

- Sample sizes n=50, 200, 1000.

- For each sample size, we simulate 1000 samples, which means there are 1000 points in each QQ plot.

- For n=1000, we generate QQ plots when null hypothesis is true and when alternative hypothesis is true.

From the four QQ plots, we can conclude that when the null hypothesis is true and the sample size is large enough ($n = 1000$), the limiting distribution of $-2 \log ELR(\theta)$ is truly a $\chi^2_{(1)}$. However, when the alternative hypothesis is true, the limiting distribution of $-2 \log ELR(\theta)$ is far from $\chi^2_{(1)}$.

```
### R codes of one example of Simulation 3 ###
### QQ plot when null hypothesis is true and g=g1 ###
### 1000 samples with sample size 1000 ###
time1<-proc.time()
myqqplot1(1000,1000,g1,1-exp(-0.5),
"Null Hypothesis is true with n=1000")
proc.time()-time1
### The time of simulation (in seconds). ###
972.596
```

6.5 Simulation 4

We repeat our QQ plots using another $g(t, \Lambda(t))$. The information of the simulated data is as follows.

- $F_0(t) = 1 - e^{-t}$, $G_0(t) = 1 - e^{-0.35t}$ denote lifetimes and censoring times respectively.

- $g_2 = g(t, \Lambda(t)) = e^{-t} e^{-\Lambda(t)} = e^{-(t+\Lambda(t))}$.

- Approximately 26% of the data are censored.
- Sample sizes n=50, 200, 1000.
- For each sample size, we simulate 1000 samples, which means there are 1000 points in each QQ plot.
- For n=1000, we generate QQ plots when null hypothesis is true and when alternative hypothesis is true.

From these QQ plots, we confirm our theorem again.

```
### R codes of one example of Simulation 4 ###
### QQ plot when null hypothesis is true and g=g2 ###
### 1000 samples with sample size 1000 ###
time1<-proc.time()
myqqplot1(1000,1000,g2,0.5,"Null Hypothesis is true with n=1000")
proc.time()-time1
### The time of simulation (in seconds). ###
2276.954
```

6.6 Simulation 5

The Gini's absolute mean difference (D) was discussed in Chapter 3. Although we assume the true distribution of income is continuous, we would use the discrete distribution to estimate Gini's abosulte mean difference (D). There is no ambiguity about the jumps of the discrete distribution. However, there are several choices for the integrand. We call $\hat{F}_n(t)$, $\hat{F}_n(t-)$ and $(\hat{F}_n(t) + \hat{F}_n(t-))/2$ the right continuous version, the left continuous version and the middle point version respectively, where $\hat{F}_n(t)$ is the empirical distribution function defined in (2.40). Suppose $X_1, \ldots, X_n$ are n i.i.d. random variables with continuous distribution function F and $x_1, \ldots, x_n$ is a realization of $X_1, \ldots, X_n$. Three versions of the estimates of Gini's absolute mean difference (D) are as follows corresponding to three different versions of integrands: the right continuous version, the left continuous version and the middle point version.

$$\begin{aligned}
\hat{D}_1 &= \sum_{i=1}^{n} x_i \left(2\hat{F}_n(x_i) - 1\right) \Delta\hat{F}_n(x_i) \\
\hat{D}_2 &= \sum_{i=1}^{n} x_i \left(2\hat{F}_n(x_i-) - 1\right) \Delta\hat{F}_n(x_i) \\
\hat{D}_3 &= \sum_{i=1}^{n} x_i \left(2\frac{\hat{F}_n(x_i) + \hat{F}_n(x_i-)}{2} - 1\right) \Delta\hat{F}_n(x_i)
\end{aligned} \tag{6.10}$$

In this simulation, we would like to compare the bias of three versions of Gini's absolute mean difference estimates. We conduct the comparison for different distributions with different sample sizes. We simulate our data from three distributions: $\chi^2_{(1)}$, Exp(1) and Log-normal(0,1) with various sample size $n = 20, 30, 50, 70, 100, 200, 500$. We simulate 100,000 samples for each comparison. The true Gini's absolute mean difference for Exp(1) is 1 and the true Gini's absolute mean difference for $\chi^2_{(1)}$ and Log-normal(0,1) are approximately 1.2732 and 1.7163 respectively, obtained through Monte Carlo simulation. From Table 6.3, we can see that the right continuous version has the smallest bias among three different versions of estimates of D. Therefore, we will use the right continuous version in the following simulations.

```
### R codes of one example of Simulation 5 ###
### The comparison of the bias of three versions of Gini's absolute
### mean difference when data is simulated from Exp(1) distribution
time1<-proc.time()
Bias(500,100000,0.5,"exp")
proc.time()-time1
### The time of simulation (in seconds). ###
```

n	Version	$\chi^2_{(1)}$	Exp(1)	Log-normal(0,1)
20	Right	0.01834396	0.02492097	0.03901017
	Left	-0.08167532	-0.07508201	-0.1257899
	Middle	-0.03166568	-0.02508052	-0.04338988
30	Right	0.01194683	0.009980541	0.02666222
	Left	-0.05470443	-0.03001954	-0.08326787
	Middle	-0.0213788	-0.0100195	-0.02830282
50	Right	0.007283817	0.009980541	0.01600711
	Left	-0.03271849	-0.03001954	-0.04995313
	Middle	-0.01271734	-0.0100195	-0.01697301
100	Right	0.003643136	0.004949118	0.008060075
	Left	-0.01635888	-0.01504903	-0.02491807
	Middle	-0.006357872	-0.005049953	-0.008428996
200	Right	0.00189952	0.00246537	0.003849105
	Left	-0.008101595	-0.00753364	-0.01263663
	Middle	-0.003101037	-0.002534135	-0.004393764
500	Right	0.000663111	0.000994784	0.00162947
	Left	-0.003336578	-0.003005141	-0.004965637
	Middle	-0.001336733	-0.001005178	-0.001668083

Table 6.3: Bias of Three Versions of Estimates of Gini's Absolute Mean Difference

15.226

6.7 Simulation 6

In this simulation we would draw the QQ plots for the uncensored data when the null hypothesis is true. The information of the simulated data is as follows.

- The uncensored data is simulated from $\chi^2_{(1)}$.
- $g_3 = g(t, F(t)) = 2t(2F(t) - 1)$
- Sample sizes n=200, 500, 1000
- For each sample size, we simulate 1000 samples, which means there are 1000 points in each QQ plot.

- For n=1000, we generate QQ plots when null hypothesis is true and when alternative hypothesis is true.

From these QQ plots, we confirm our Theorem 4.2.7 in Chapter 4.

```
### R codes of one example of Simulation 6 ###
### QQ plot when null hypothesis is true and g=ginimdf ###
### 1000 samples with sample size 1000 ###
time1<-proc.time()
myqqplot2(1000,1000,ginimdf,0.6366,1e-8,
"Null Hypothesis is true with n=1000")
proc.time()-time1
### The time of simulation (in seconds). ###
 22.515
```

6.8 Simulation 7

In this section, a simulation study of the Gini's absolute mean difference (D) will be presented. Before that, we will have a brief discussion of the history of the variance estimation of Gini index, which is also called the Gini's relative mean difference. See Langel and Tille (2011) for a detailed discussion of this topic.

Before 1980s, a very limited number of papers have focused on the variance estimation of the Gini index. Nair (1936) computes the exact variance of Gini's absolute mean difference for the first time. Nevertheless, the expression of variance he have obtained is particularly cumbersome. Lomnicki(1952) and Glasser(1962) approximate Nair's expression and propose simpler variance estimators.

One of the very first results of the variance of the Gini index is obtained by Hoeffding (1948), who expresses the Gini index as a function of two U-statistics. Seminal works on the variance estimation of the Gini index are attributed to Sandstrom et al. (1985,1988), who discuss four variance estimators for the Gini index. For more references on the variance estimations of the Gini index, see Yitzhaki (1991), Karagiannis and Kovacevic (2000) and many others. However, the previous authors have never studied the confidence intervals of the Gini index, except Sandstrom et al. (1989) who briefly mention 95% normal approximation confidence intervals based on three variance estimators. The problem of the confidence intervals based on normal approximation is that it is not guaranteed that the confidence interval will be within the valid domain of parameters. Some kinds of transformations are needed. However, we do not always know which transformation to use or which transformation is the best. If this is the case, empirical likelihood method may be used, which does not need the transformation and the variance estimation. See Qin et al. (2010) and Peng (2011) for discussions of the confidence intervals of the Gini index using empirical likelihood method.

Next, we would introduce two empirical likelihood ratio methods for Gini index and Gini's absolute mean difference. To begin with, we point out that the statistical functional of the Gini's absolute mean difference $T(F) = \int 2t(2F(t) - 1)dF(t)$ is Hadamard differentiable. First, it is obvious that $g(t, F(t)) = 2t(2F(t) - 1)$ is left continuous with respect to t and twice differentiable with respect to $F(t)$. Second, $g(t, F(t))$ is bounded by $A(t) = 2t$ and $A(t)$ is integrable with respect to any distribution function since we assume that the all income distribution functions have the finite mean. Third, $h(t, F(t)) = 2t$ is bounded by $A(t) = 2t$ and $A(t)$ is integrable with respect to $F_0(t)$ since $F_0(t)$ has the finite mean.

Qin et al. (2010) prove the following theorem. Let $(y_1, \ldots, y_n)$ be an i.i.d. sample from distribution function F and $\hat{F}_n$ be the empirical distribution function. The simple plug-in moment estimator of the Gini index (G) is given by

$$\hat{G} = \frac{1}{\hat{\mu}}\frac{1}{n}\sum_{i=1}^{n}\left(y_i\left(2\hat{F}_n(y_i)-1\right)\right) \tag{6.11}$$

where $\hat{\mu} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

The log-EL ratio statistic for $\theta = G$ is given by

$$R(\theta) = \sum_{i=1}^{n}\log\{n\tilde{p}_i(\theta)\} \tag{6.12}$$

where $\tilde{p}_1(\theta), \ldots, \tilde{p}_n(\theta)$ maximize the log-EL function $l(\mathbf{p}) = \sum_{i=1}^{n}\log(p_i)$ subject to the following set of constraints:

$$p_i > 0,\ \sum_{i=1}^{n} p_i = 1,\ \sum_{i=1}^{n} p_i\left(\left(2\hat{F}_n(y_i)-1\right)y_i - \theta y_i\right) = 0 \tag{6.13}$$

Theorem 6.8.1 (Qin et al.(2011)) *Suppose that $0 < E(Y^3) < \infty$. Then, as $n \to \infty$,*

$$-2R(\theta) \xrightarrow{D} \frac{\sigma_3^2}{\sigma_2^2}\chi^2_{(1)} \tag{6.14}$$

where Y is a random variable denoting the income of a population with distribution function $F(y)$, $\sigma_2^2 = Var(2YF(Y) - (\theta+1)Y)$, $\sigma_3^2 = Var(2h_1(Y) - (\theta+1)Y)$ and $h_1(y) = yF(y) + \int_y^{+\infty} x dF(x)$.

By Theorem 6.8.1, the $100(1-\alpha)\%$ confidence interval based on EL method can be constructed as

$$\{\theta | -2R(\theta) \le \hat{k}^{-1}\chi^2_{\alpha,1}\} \tag{6.15}$$

where $\chi^2_{\alpha,1}$ is the $(1-\alpha)$th quantile of $\chi^2_{(1)}$ and $\hat{k}$ is given by $\hat{k} = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_3^2}$ where

$$\hat{\sigma}_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(u_{2i} - \bar{u}_2\right)^2 \tag{6.16}$$

with

$$u_{2i} = 2y_i\hat{F}_n(y_i) - \left(\hat{G}+1\right)y_i, \quad \bar{u}_2 = \frac{1}{n}\sum_{j=1}^{n}u_{2j}, \tag{6.17}$$

and

$$\hat{\sigma}_3^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(u_{1i} - \bar{u}_1\right)^2 \tag{6.18}$$

with

$$u_{1i} = 2\hat{h}_1(y_i) - \left(\hat{G}+1\right)y_i, \quad \bar{u}_1 = \frac{1}{n}\sum_{i=1}^{n}u_{1i}, \tag{6.19}$$

and

$$\hat{h}_1(y) = y\hat{F}_n(y) + \frac{1}{n}\sum_{j=1}^{n}y_j I\left(y_j \geq y\right) \tag{6.20}$$

They also provide a bootstrap-calibration method to avoid the estimation of the scale parameter k.

Peng (2011) proves the following theorem. Let $(X_1, \ldots, X_n)$ be n i.i.d. random variables with distribution function F. Let m be the integer part of $\frac{n}{2}$, and define $Y_i = \frac{X_i + X_{m+i}}{2}$ and $Z_i = \min(X_i, X_{m+i})$ for $i = 1, \ldots, m$. It is easy to check that

$$E\left[Y_i - Z_i - Y_i G\right] = 0 \tag{6.21}$$

where G is the Gini index.

Peng (2011) defines an empirical likelihood ratio function for $\theta = G$ as follows.

$$L_1(\theta) = \prod_{i=1}^{m}\left(mp_i\right) \tag{6.22}$$

where $p_i = \Delta F(X_i) = F(X_i) - F(X_i-)$, $i = 1, \ldots, m$.

He maximizes the above empirical likelihood ratio function subject to the following constraints:

$$p_i \geq 0, i = 1, \ldots, m, \quad \sum_{i=1}^{m} p_i = 1, \quad \sum_{i=1}^{m} p_i \left(Y_i - Z_i - Y_i\theta\right) = 0 \tag{6.23}$$

By the Lagrange multiplier method, the maximum of the log-empirical likelihood ratio function is as follows.

$$l_1(\theta) = -2\log\left(\sup L_1(\theta)\right) = 2\sum_{i=1}^{m} \log\left(1 + \lambda\left(Y_i - Z_i - Y_i\theta\right)\right) \tag{6.24}$$

where λ satisfies

$$\frac{1}{m}\sum_{i=1}^{m} \frac{Y_i - Z_i - Y_i\theta}{1 + \lambda\left(Y_i - Z_i - Y_i\theta\right)} = 0 \tag{6.25}$$

The following theorem directly follows Theorem 2 of Qin and Lawless (1994).

Theorem 6.8.2 (Peng (2011)) *Assume $E(X_1^3) < \infty$. Then*

$$l_1(\theta) \xrightarrow{D} \chi^2_{(1)}, \ \textit{as } n \to \infty \tag{6.26}$$

Based on Theorem 6.8.2, a $100(1-\alpha)\%$ confidence interval can be constructed as follows.

$$\{\theta | l_1(\theta) \leq \chi^2_{\alpha,1}\} \tag{6.27}$$

where $\chi^2_{\alpha,1}$ is the $(1-\alpha)$th quantile of $\chi^2_{(1)}$.

The problem of Peng (2011)'s approach is explicit. Since he only uses half of the data, his estimator might be unbiased, which means the coverage probability of his confidence intervals might be good, but the variance of his estimator shall be large. We expect that the average length of Peng's confidence intervals shall be larger than ours.

Both Qin et al.'s plug-in method and our Hadamard derivative method try to linearize the complex, non-linear constraint of the Gini index. However, the plug-in method only captures part of the linear component while our derivative method captures all the linear part of the non-linear constraint. To see this, we use the non-linear statistical functional of the Gini's absolute mean difference $T(F) = \int g(t, F(t))dF(t) = \int 2t(2F(t)-1)dF(t)$ as an example. Qin et al. plug the empirical distribution function $\hat{F}_n(t)$ in $g(t, F(t))$ and linearize the non-linear statistical functional of the Gini's absolute mean difference as $\int g(t, \hat{F}_n(t))dF(t) = \int 2t(2\hat{F}_n(t) - 1)dF(t)$. From Theorem 4.2.3, it can been shown that the Hadamard derivative of $T(F) = \int g(t, F(t))dF(t) = \int 2t(2F(t) - 1)dF(t)$ at $\hat{F}_n(t)$ is $T'_{\hat{F}_n}(F(t)) = \int (g(t, \hat{F}_n(t)) + \tilde{h}(t))dF(t)$, where $\tilde{h}(t) = \int_t^{+\infty} h(s, \hat{F}_n(s))d\hat{F}_n(s)$ and $h = \frac{\partial g}{\partial F}$. Qin et al.'s plug-in linearization is only the first part of our Hadamard derivative linearization and miss the second part. The partial linearization of Qin et al.'s plug-in method is the reason that their empirical likelihood ratio converges to a weighted chi-square distribution under the null hypothesis, rather than a chi-square distribution. Our empirical likelihood ratio has a limiting distribution of a chi-square distribution under the null hypothesis also confirms that we capture all the linear component.

Since the limiting distribution of Qin et al.'s empirical likelihood ratio is a weighted chi-square distribution, they have to come up with a consistent estimate of the weight. This task becomes so complicated when data are censored. Further more if we are dealing with a limiting distribution of a chi-square distribution with degrees of freedom above one, then Qin et al.'s method not only is too complicated, but also loses power compared to ours, while our method remains a clean chi-square distribution. See Zhou (2015) chapter 7 for details.

In the following, we compare the coverage probability and average length of the confidence intervals of the Gini's absolute mean difference (D) based on our method, Qin et al. (2010)'s method and Peng (2011)'s method. Qin et al.'s method and Peng's method are just discussed above. Our method is mentioned in Simulation 1 of Chapter 6 (6.9), where ELR function is in terms of distribution functions and defined in (4.56) in Chapter 4. Our method uses the iterative algorithm in section 6.1 of this chapter to compute the maximum of the ELR function, but uses the empirical distribution function as the initial plug-in value for $F(t)$ in $g(t, F(t))$.

The empirical likelihood confidence intervals have the under coverage problem. Owen has already discussed this issue in his book *Empirical Likelihood*. One possible way to correct this problem is to use the F quantile instead of the chi-square quantile. $F_{1-\alpha,1,n-2}$, which is the $(1-\alpha)$th percentile of $F(1, n-2)$ distribution, is larger than $\chi^2_{1,1-\alpha}$, which is the $(1-\alpha)$th percentile of a chi-square distribution with one degree of freedom. And the difference between these two quantiles decreases with respect to sample size n. Therefore, using F quantile instead of chi-square quantile will improve the coverage probability of empirical likelihood confidence intervals, especially for the small sample size. The F quantile calibrated confidence interval based on our method is

$$\{\theta \mid \theta \;\; s.t. \;\; -2\log ELR(\theta) \leq F_{1-\alpha,1,n-2}\} \tag{6.28}$$

We simulate our data from three different distributions: $\chi^2_{(1)}$, EXP(1) and Log-normal(0,1) with various sample size $n = 50, 70, 100, 300, 500$. The nominal level of the confidence intervals is 0.95. For coverage probability, we simulate 5000 samples for each comparison and for average length, we simulate 100 confidence intervals for each comparison.

n	Method	$\chi^2_{(1)}$	Exp(1)	Log-normal(0,1)
50	Our	0.9154	0.9138	0.8844
	Qin et al.	0.9072	0.913	0.8732
	Peng	0.9058	0.9228	0.8804
70	Our	0.9236	0.9244	0.8916
	Qin et al.	0.9198	0.931	0.8892
	Peng	0.9256	0.9276	0.9004
100	Our	0.933	0.9248	0.8982
	Qin et al.	0.9294	0.9328	0.897
	Peng	0.9322	0.9356	0.9054
300	Our	0.939	0.927	0.9214
	Qin et al.	0.9398	0.9372	0.9256
	Peng	0.9422	0.943	0.9266
500	Our	0.943	0.9344	0.9238
	Qin et al.	0.943	0.9454	0.9304
	Peng	0.9454	0.9466	0.9366

Table 6.4: Coverage Probability Comparison

Table 6.4 shows that the coverage probability of Peng's confidence intervals is superior to our confidence intervals after F quantile calibration and Qin et al.'s in all three distributions with all sample sizes. The coverage probability of our confidence intervals after F quantile calibration is better than Qin et al.'s when sample size is small while the coverage probability of Qin et al.'s confidence intervals is better than our confidence intervals after F quantile calibration when sample size is large. All three methods have the under cover problem with Log-normal(0,1) distribution.

Table 6.5, as our expectation, shows that Peng's confidence intervals are longer than our confidence intervals after F quantile calibration and Qin et al.'s in all three distributions with all sample sizes. Our confidence intervals after F quantile calibration are longer than Qin et al.'s when sample size is small and shorter than Qin et al.'s when sample size is large.

n	Method	$\chi^2_{(1)}$	Exp(1)	Log-normal(0,1)
50	Our	0.539	0.318585	0.81725
	Qin et al.	0.515865	0.310755	0.787195
	Peng	0.57047	0.37932	0.86949
70	Our	0.444575	0.27941	0.71313
	Qin et al.	0.429755	0.27634	0.69494
	Peng	0.493795	0.33708	0.76553
100	Our	0.37721	0.23178	0.596515
	Qin et al.	0.369265	0.232405	0.5882
	Peng	0.424295	0.28252	0.655475
300	Our	0.21334	0.12682	0.35427
	Qin et al.	0.212645	0.1312	0.35436
	Peng	0.246775	0.161085	0.39439
500	Our	0.166895	0.09488	0.28761
	Qin et al.	0.16747	0.09922	0.28908
	Peng	0.19265	0.12235	0.31904

Table 6.5: Average Length of Confidence Intervals Comparison

```
### R codes of one example of Simulation 7: Coverage Probability

### The comparison of the coverage probabilities of

### three empirical likelihood confidence intervals.

### 5000 samples with sample size 500

### when data is simulated from chisq(1)

time1<-proc.time()

CP_Comp(500,5000,0.6366,ginimdf,h1,1e-8,0.001,"chisq",cali=T)

proc.time()-time1

### The time of simulation (in seconds). ###

 56.994


### R codes of one example of Simulation 7: Average length comparison

### The comparison of the average length of

### three empirical likelihood confidence intervals.

### 100 samples with sample size 500

### and data is simulated from chisq(1)
```

Year	$\hat{G}$	$\hat{D}_1$($)	Our($)	Qin($)	Our(P-value)	Qin(P-value)
2000	0.5357	9920	(8640,11370)	(8610,11356)	0.9062	0.9533
2001	0.5322	10138	(8852,11586)	(8822,11578)	0.8389	0.9209
2002	0.5294	10443	(9146,11892)	(9118,11888)	0.5119	0.7531
2003	0.5280	10884	(9548,12376)	(9520,12372)	0.1977	0.5402

Table 6.6: The Nominal 95% Empirical Likelihood Confidence Intervals of the Gini's Absolute Mean Difference Based on the Real GDP Per Capita in Constant Dollars Expressed in International Prices (Base Year 2000)

```
time1<-proc.time()

AVL_Comp(500,100,ginimdf,h1,0.6366,1e-8,0.05,0.4,0.8,0.0005,0.1,0.0005,

"chisq",cali=T)

proc.time()-time1

### The time of simulation (in seconds). ###

 1328.50
```

6.9 Real Data Analysis

In this section, we apply our method and Qin et al.'s method to the real GDP per capita in constant dollars expressed in international prices from 2000 to 2003 (2000 as the base year). Therefore, the Gini index is a measure of the dispersion of consumption across 182 countries of which data are available. These data sets are from the Penn World Tables(Summers & Heston (1995)).

In Table 6.6, we report the nonparametric estimator $\hat{G}$ and $\hat{D}_1$ given in (6.11) and (6.10) respectively and the nominal 95% confidence intervals of the Gini's absolute mean difference (D) and P-values based on our method and Qin et al.'s method. The P-values are corresponding to the null hypothesis $H_0 : D = \$10000$. This table indicates that the interval lengths of the confidence intervals based on our method are shorter than those based on Qin et al.'s method.

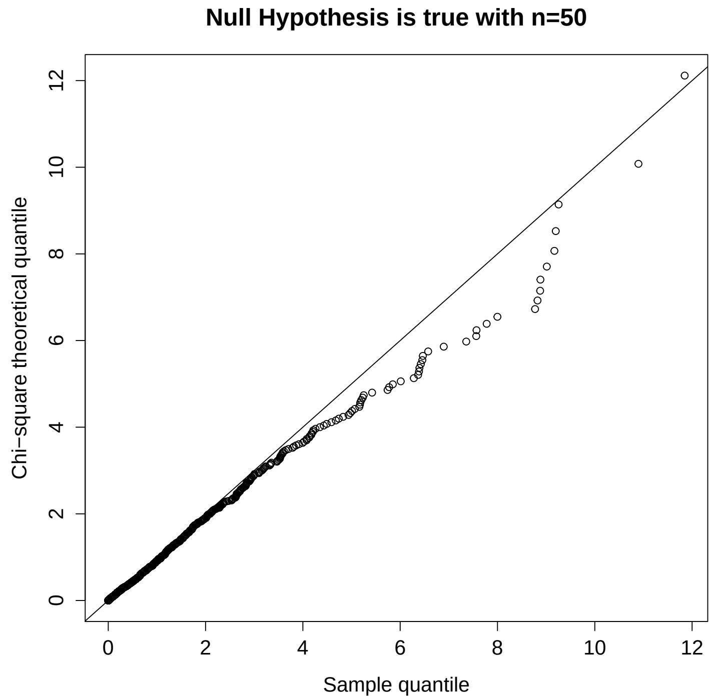

Figure 6.2: Simulation 3, $g(t, \Lambda(t)) = I[t \leq 0.5]e^{-\Lambda(t)}$ and $n = 50$

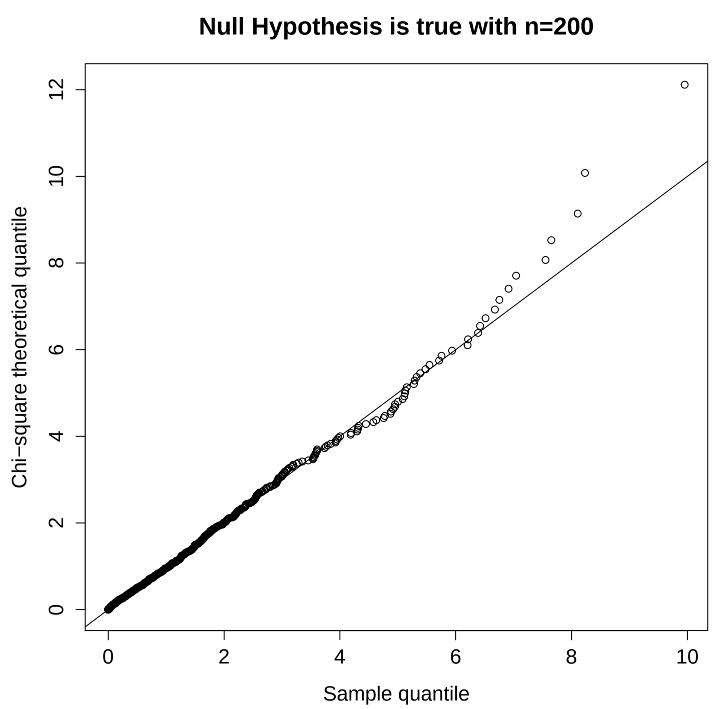

Figure 6.3: Simulation 3, $g(t, \Lambda(t)) = I[t \leq 0.5]e^{-\Lambda(t)}$ and $n = 200$

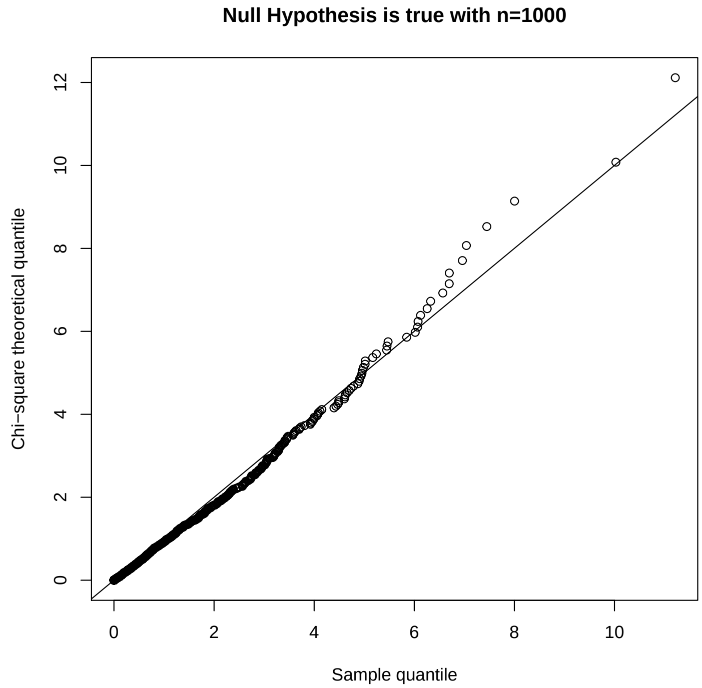

Figure 6.4: Simulation 3, $g(t, \Lambda(t)) = I[t \leq 0.5]e^{-\Lambda(t)}$ and $n = 1000$ when Null Hypothesis is True

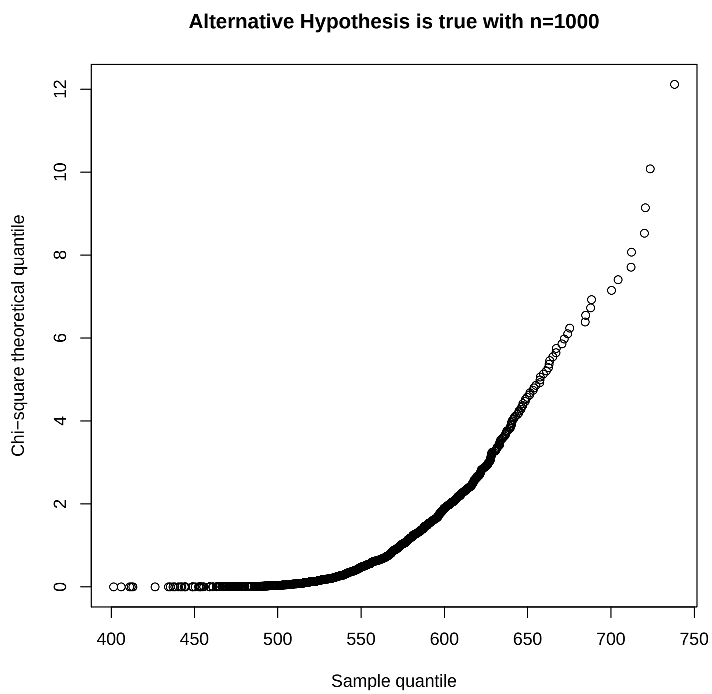

Figure 6.5: Simulation 3, $g(t, \Lambda(t)) = I[t \leq 0.5]e^{-\Lambda(t)}$ and $n = 1000$ when Alternative Hypothesis is True

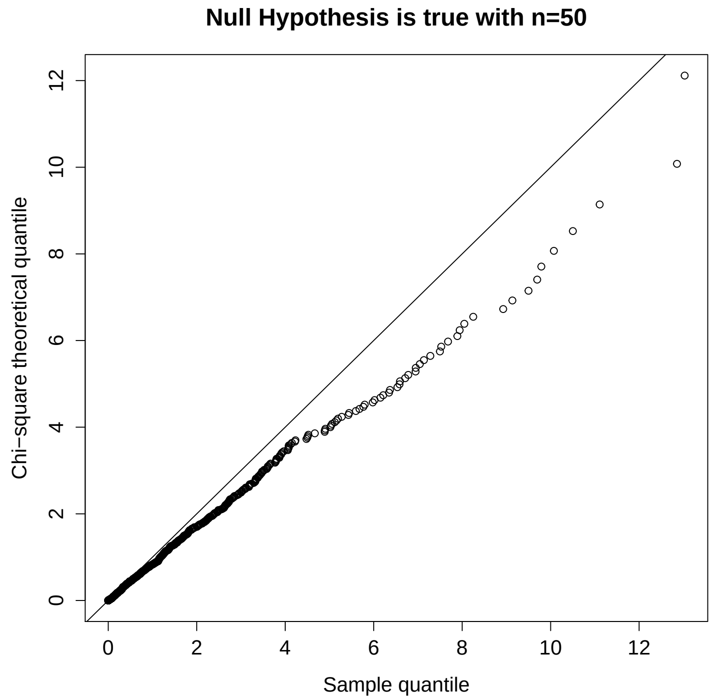

Figure 6.6: Simulation 4, $g(t, \Lambda(t)) = e^{-(t+\Lambda(t))}$ and $n = 50$

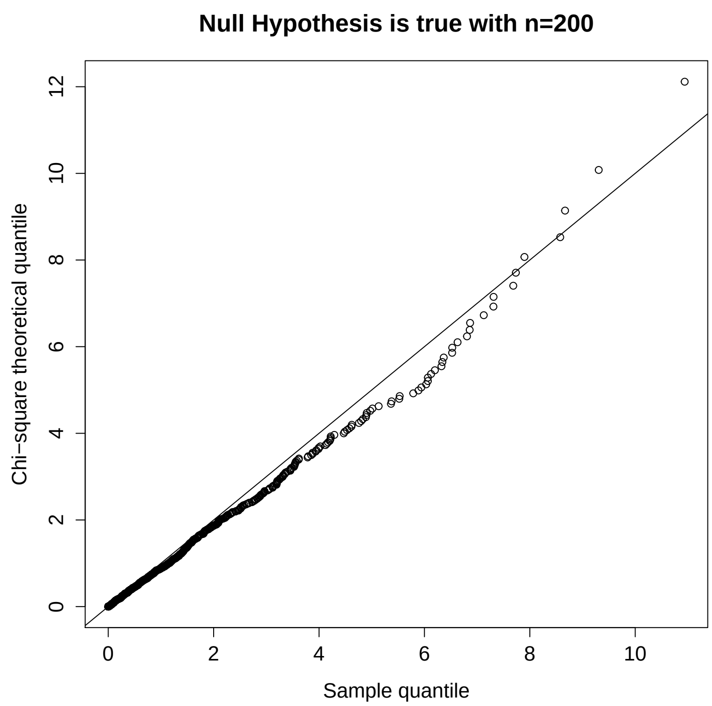

Figure 6.7: Simulation 4, $g(t, \Lambda(t)) = e^{-(t+\Lambda(t))}$ and $n = 200$

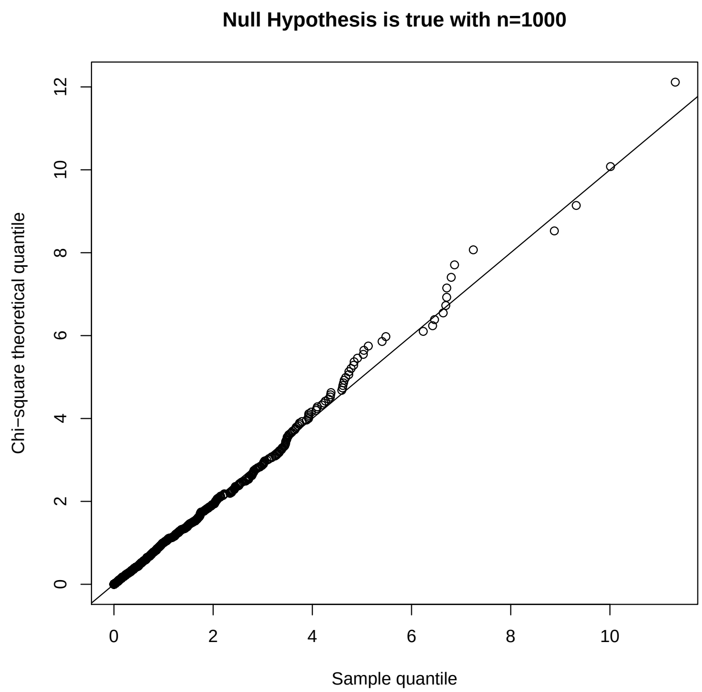

Figure 6.8: Simulation 4, $g(t, \Lambda(t)) = e^{-(t+\Lambda(t))}$ and $n = 1000$ when Null Hypothesis is True

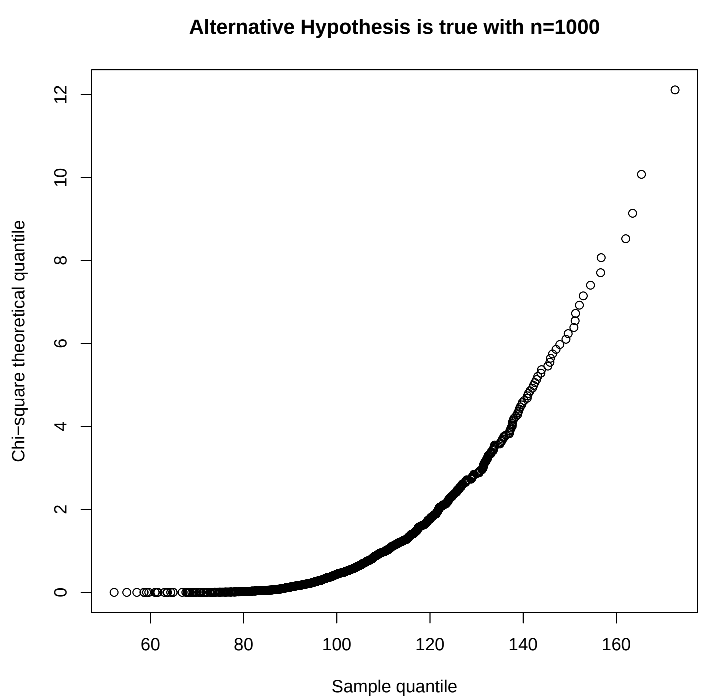

Figure 6.9: Simulation 4, $g(t, \Lambda(t)) = e^{-(t+\Lambda(t))}$ and $n = 1000$ when Alternative Hypothesis is True

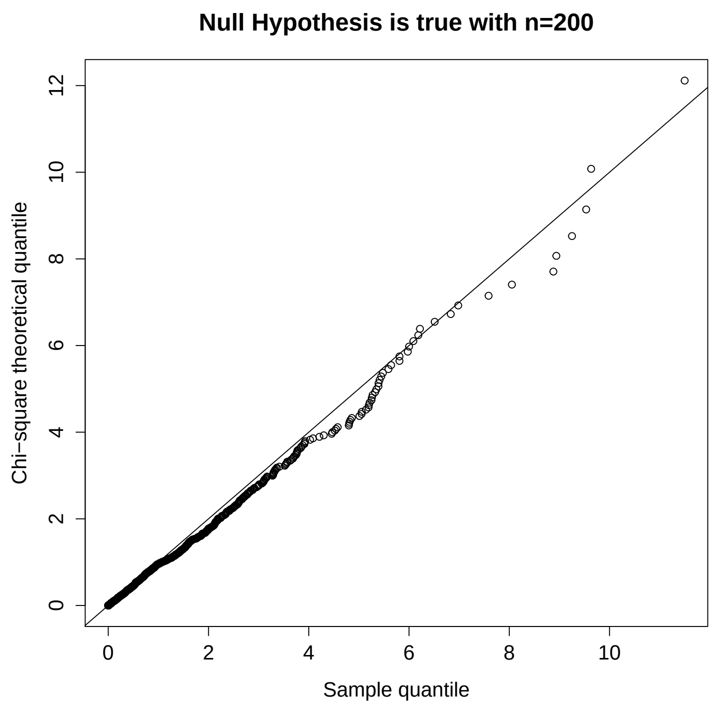

Figure 6.10: Simulation 6, $g(t, F(t)) = 2t(2F(t) - 1)$ and $n = 200$

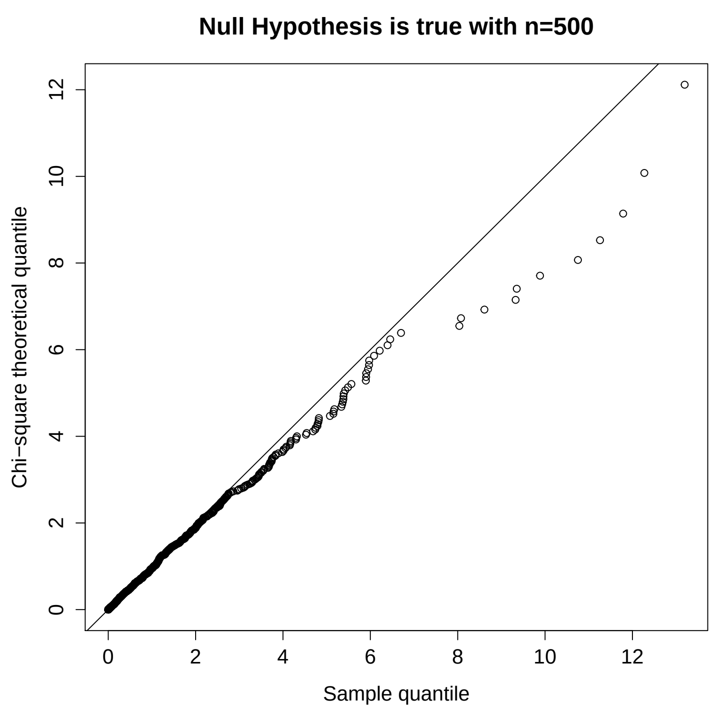

Figure 6.11: Simulation 6, $g(t, F(t)) = 2t(2F(t) - 1)$ and $n = 500$

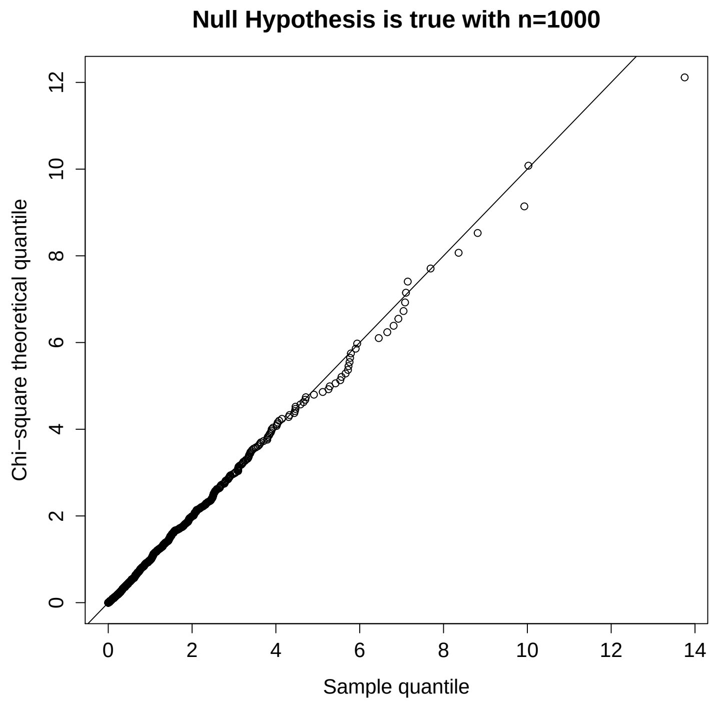

Figure 6.12: Simulation 6, $g(t, F(t)) = 2t(2F(t) - 1)$ and $n = 1000$ when Null Hypothesis is True

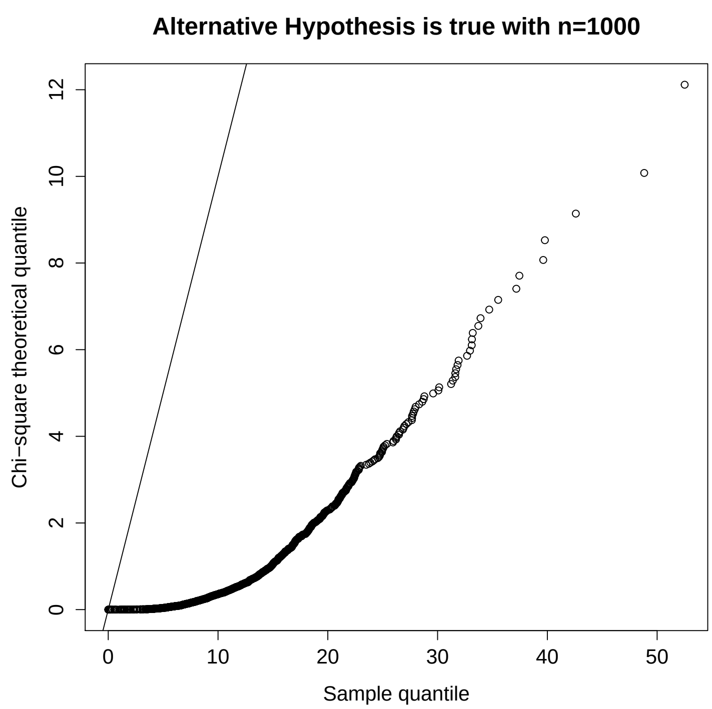

Figure 6.13: Simulation 6, $g(t, F(t)) = 2t(2F(t) - 1)$ and $n = 1000$ when Alternative Hypothesis is True

Chapter 7 Discussion and Future Questions

We have already mentioned in Chapter 3 that it is difficult to generalize Owen's result to the right censored data setting since there is no explicit maximum of distribution function of Lagrange multiplier method. However, it is possible to solve the problem computationally. There are several methods available to compute the empirical likelihood ratio with distribution-type constraint for right censored data.

Chen and Zhou (2007) propose to use a sequential quadratic programming (SQP) method to compute the empirical likelihood ratio with mean-type constraint for right censored data. The SQP is a nonlinear programming method. See Nocedal and Wright (1999) for more details. Instead of applying the SQP method directly, Chen and Zhou introduce several auxiliary variables, which makes the matrix $\boldsymbol{G}$ diagonal. This technique simplifies the computation of SQP tremendously.

Zhou (2005) proposes an EM algorithm to compute the empirical likelihood ratio. He also compares the EM algorithm to the SQP method and concludes that EM algorithm is superior to the SQP method.

For the same distribution-type constraint,

$$\int g(t)dF(t) = \theta_0 \tag{7.1}$$

Zhou and Yang (2015) find a recursive formula to compute the empirical likelihood, which outperforms the SQP method and EM algorithm. The recursive formula of the jump of the distribution function is as follows.

$$w_i = \Delta F(T_i) = \frac{\delta_i}{n - \lambda g(T_i) - \sum_{j:\delta_j=0} \frac{I[T_j<T_i]}{S_j}} \tag{7.2}$$

where $S_j = \sum_{T_i<T_j} w_i = 1 - \sum_{T_i \le T_j} w_i$.

It is difficult to apply the SQP method or EM algorithm to the computation of empirical likelihood subject to the constraint of the following form.

$$\int g(t, \Lambda(t)) d\Lambda(t) = \theta_0 \tag{7.3}$$

The recursive method may be applied under some specific conditions i.e. $g(t, \Lambda(t)) = te^{-\Lambda(t)}$.

To be more clear, the AL of a sample $(T_1, \delta_1), \ldots, (T_n, \delta_n)$ as defined before is

$$AL = \prod_{i=1}^{n} w_i^{\delta_i} \exp\left\{-\sum_{j=1}^{i} w_j\right\} \tag{7.4}$$

where $w_i = \Delta\Lambda(T_i)$.

The constraint (7.3) can be rewritten in the discrete form as follows.

$$\sum_{i=1}^{n-1} \delta_i g(T_i, \sum_{j=1}^{i} w_j) w_i + \delta_n g(T_n, \sum_{j=1}^{n} w_j) = \theta_0 \tag{7.5}$$

In order to apply the Lagrange multiplier method, we form the target function

$$\begin{aligned} G = & \sum_{i=1}^{n} \delta_i \log w_i - \sum_{i=1}^{n} \sum_{j=1}^{i} w_j \\ & + n\lambda \left(\theta_0 - \sum_{i=1}^{n-1} \delta_i g(T_i, \sum_{j=1}^{i} w_j) w_i - \delta_n g(T_n, \sum_{j=1}^{n} w_j)\right) \end{aligned} \tag{7.6}$$

Note that $w_n = \delta_n$; taking the derivative with respect to $w_i, i = 1, \dots, n-1$ and equaling them to 0, we have

$$\begin{aligned}\frac{\partial G}{\partial w_i} &= \frac{\delta_i}{w_i} - (n-i+1) - n\lambda \left(\sum_{l=i}^{n-1} \delta_l h(T_l, \sum_{j=1}^{l} w_j) w_l + g(T_i, \sum_{j=1}^{i} w_j) + \delta_n h(T_n, \sum_{j=1}^{n} w_j) \right) \\ &= \frac{\delta_i}{w_i} - (n-i+1) - n\lambda \sum_{l=i}^{n} \delta_l h(T_l, \sum_{j=1}^{l} w_j) w_l - n\lambda g(T_i, \sum_{j=1}^{i} w_j)\end{aligned} \tag{7.7}$$

where $h = \frac{\partial g}{\partial \Lambda}$.

The jump of cumulative hazard function Λ at time T_i is

$$w_i = \frac{\delta_i}{(n-i+1) + n\lambda \sum_{l=i}^{n} \delta_l h(T_l, \sum_{j=1}^{l} w_j) w_l + n\lambda g(T_i, \sum_{j=1}^{i} w_j)} \tag{7.8}$$

In some specific situations like $g(t, \Lambda(t)) = te^{-\Lambda(t)}$, $h(t, \Lambda(t)) = -te^{-\Lambda(t)} = -g(t, \Lambda(t))$. Then we have

$$\sum_{i=1}^{n-1} \delta_i h(T_i, \sum_{j=1}^{i} w_j) w_i + \delta_n h(T_n, \sum_{j=1}^{n} w_j) = -\theta_0 \tag{7.9}$$

so

$$\sum_{l=i}^{n} \delta_l h(T_l, \sum_{j=1}^{l} w_j) w_l = -\theta_0 - \sum_{l=1}^{i-1} \delta_l h(T_l, \sum_{j=1}^{l} w_j) w_l \tag{7.10}$$

Plugging this into (7.8), we obtain a recursive equation for w_i,

$$w_i = \frac{\delta_i}{(n-i+1) + n\lambda \left(-\theta_0 - \sum_{l=1}^{i-1} \delta_l h(T_l, \sum_{j=1}^{l} w_j) w_l + g(T_i, \sum_{j=1}^{i-1} w_j + w_i) \right)} \tag{7.11}$$

Suppose we have the λ fixed first, when $i = 1$, w_1 is the solution of the following nonlinear equation.

$$w_1 = \frac{\delta_1}{n + n\lambda g(T_1, w_1)} \tag{7.12}$$

As long as we have $w_i, i = 1, \ldots, k$, w_{k+1} is the solution of the following nonlinear equation.

$$w_{k+1} = \frac{\delta_{k+1}}{(n-(k+1)+1) + n\lambda\left(-\theta_0 - \sum_{l=1}^{k} \delta_l h(T_l, \sum_{j=1}^{l} w_j) w_l + g(T_i, \sum_{j=1}^{k} w_j + w_{k+1})\right)} \tag{7.13}$$

Once we have all w_i, we plug them into (7.5). The constraint may or may not be θ_0. If the constraint does not equal to θ_0, we actually get the maximum under that constraint. We can change the value of λ to get another set of w_i and plug them into (7.5) until the constraint equals to θ_0.

Chapter 8 Appendix

8.1 Additional Lemmas of Chapter 3

Theorem 8.1.1 *Let f_n be a sequence of real-valued measurable functions on a measure space (S, Σ, μ). Suppose that the sequence converges pointwisely to a function f and is dominated by some integrable function g in the sense that*

$$|f_n(x)| \leq g(x) \tag{8.1}$$

for all numbers n in the index set of the sequence and all points $x \in S$.Then f is integrable and

$$\lim_{n\to\infty} \int_S |f_n - f| d\mu = 0 \tag{8.2}$$

which also implies

$$\lim_{n\to\infty} \int_S f_n d\mu = \int_S f d\mu \tag{8.3}$$

Lemma 8.1.2 (Chow and Teicher (1980)) *For any random variable Y, if $E|Y|^k < \infty$, then for an i.i.d. sample $Y_1, Y_2, \ldots, Y_n$ that has the same distribution as Y, we have*

$$\max_{1\leq i\leq n} |Y_i| = o\left(n^{\frac{1}{k}}\right) \quad a.s. \tag{8.4}$$

Proof See Chow and Teicher (1980, p. 131, problem No. 8). ∎

Let

$$M_n = \max_{1\leq i\leq n} |Z_i| \tag{8.5}$$

where $Z_i = \frac{\delta_i\left(g(T_i,\Lambda_0(T_i))+\tilde{h}(T_i)\right)}{\frac{n-i+1}{n}}$.

To prove that $M_n = o_p(n^{\frac{1}{2}})$, we need the following lemma.

Lemma 8.1.3 (Pan and Zhou (2002)) *Let $(T_1, \delta_1), \ldots, (T_n, \delta_n)$ be n i.i.d. pairs of random variables, where each (T_i, δ_i) is defined by (3.42). Let also $T_n^* = \max_{1 \le i \le n} T_i$. If $\int h^2(x) d\Lambda_0(x) < \infty$, then*

$$\max_{1 \le i \le n} \frac{\delta_i |h(T_i)|}{\sqrt{(1 - F_0(T_i))(1 - G_0(T_i))}} = o(n^{\frac{1}{2}}) \quad a.s. \quad and \quad \delta_n^* h(T_n^*) = o_p(1), \tag{8.6}$$

where δ_n^* is the indicator function corresponding to T_n^*.

Proof See Pan and Zhou (2002) Lemma A2. ∎

Now we have

$$\begin{aligned} \max_{1 \le i \le n} |Z_i| &= \max_{1 \le i \le n} \left| \frac{\delta_i \left(g(T_i, \Lambda_0(T_i)) + \tilde{h}(T_i) \right)}{\frac{n-i+1}{n}} \right| \\ &\le \max_{1 \le i \le n} \left| \frac{\delta_i \left(g(T_i, \Lambda_0(T_i)) + \tilde{h}(T_i) \right)}{(1 - F_0(T_i))(1 - G_0(T_i))} \right| \times \max_{1 \le i \le n} \left| \frac{(1 - F_0(T_i))(1 - G_0(T_i))}{\frac{n-i+1}{n}} \right| \end{aligned} \tag{8.7}$$

Use Lemma (8.1.3) and choose

$$h(x) = \left| \frac{g(x, \Lambda_0(x)) + \tilde{h}(x)}{\sqrt{(1 - F_0(x))(1 - G_0(x))}} \right| \tag{8.8}$$

If we assume that

$$\int \frac{\left(g(x, \Lambda_0(x)) + \tilde{h}(x) \right)^2}{(1 - F_0(x))(1 - G_0(x))} d\Lambda_0(x) < \infty \tag{8.9}$$

we have

$$\max_{1 \le i \le n} \frac{|\delta_i| |g(T_i, \Lambda_0(T_i)) + \tilde{h}(T_i)|}{(1 - F_0(T_i))(1 - G_0(T_i))} = o(n^{\frac{1}{2}}) \tag{8.10}$$

It is obvious that

$$\max_{1\le i\le n}\left|\frac{(1-F_0(T_i))\,(1-G_0(T_i))}{\frac{n-i+1}{n}}\right| = O_p(1) \tag{8.11}$$

then we have

$$(8.7) \le o(n^{\frac{1}{2}}) \times O_p(1) = o_p(n^{\frac{1}{2}}) \tag{8.12}$$

so we have

$$M_n = \max_{1\le i\le n} |Z_i| = o_p(n^{\frac{1}{2}}) \tag{8.13}$$

∎

Lemma 8.1.4 *Under the assumption of Lemma 3.2.5, we have, for Z_i defined in Lemma 3.2.3,*

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i - \hat{\theta} - \theta_0\right) \xrightarrow{D} N(0, \sigma^2_{\Lambda_0}) \tag{8.14}$$

where $\sigma^2_{\Lambda_0} = \int \frac{\left(g(x,\Lambda_0(x))+\tilde{h}(x)\right)^2}{(1-F_0(x))(1-G_0(x))} d\Lambda_0(x)$, $\hat{\theta} = \int \tilde{h}(t) d\Lambda_0(t)$ *and* $\theta_0 = \int g(x, \Lambda_0(x)) d\Lambda_0(x)$

Proof The summation can be written as

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} Z_i - \hat{\theta} - \theta_0 \\
&= \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i\left(g(T_i,\Lambda_0(T_i)) + \tilde{h}(T_i)\right)}{\frac{n-i+1}{n}} - \hat{\theta} - \theta_0 \\
&= \sum_{i=1}^{n}\left(g(T_i,\Lambda_0(T_i)) + \tilde{h}(T_i)\right)\Delta\hat{\Lambda}_{NA}(T_i) - \hat{\theta} - \theta_0 \\
&= \int\left(g(t,\Lambda_0(t)) + \tilde{h}(t)\right) d\hat{\Lambda}_{NA}(t) - \int\left(g(t,\Lambda_0(t)) + \tilde{h}(t)\right) d\Lambda_0(t) \\
&= \int\left(g(t,\Lambda_0(t)) + \tilde{h}(t)\right) d\left(\hat{\Lambda}_{NA}(t) - \Lambda_0(t)\right)
\end{aligned} \tag{8.15}$$

Similar arguments to Andersen et al. (1993, Chap. 4) can be used to analyze the integral. Since $\left(g\left(t,\Lambda_0(t)\right) + \tilde{h}(t)\right)$ is left continuous, it is predictable. An application of the martingale central limit theorem will finish the proof. ∎

Lemma 8.1.5 *Under the assumption of Lemma 3.2.5, and we have, for Z_i defined in Lemma 3.2.3,*

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n} Z_i^2 \xrightarrow{P} \int \frac{\left(g(t,\Lambda_0(t))+\tilde{h}(t)\right)^2}{(1-F_0(t))\,(1-G_0(t))} d\Lambda_0(t),\ \ as\ n\ \to\infty \\ &\frac{1}{n}\sum_{i=1}^{n} \frac{Z_i^2}{n-i} \xrightarrow{P} 0,\ \ as\ n\ \to\infty\end{aligned} \tag{8.16}$$

Proof

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n} Z_i^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\delta_i\left(g(T_i,\Lambda_0(T_i))+\tilde{h}(T_i)\right)}{\frac{n-i+1}{n}}\right)^2 \\ &= \sum_{i=1}^{n}\frac{\delta_i}{n-i+1}\frac{1}{\frac{n-i+1}{n}}\left(g(T_i,\Lambda_0(T_i))+\tilde{h}(T_i)\right)^2 \\ &= \sum_{i=1}^{n}\frac{\left(g(T_i,\Lambda_0(T_i))+\tilde{h}(T_i)\right)^2}{\frac{Y(T_i)}{n}}\Delta\hat{\Lambda}_{NA}(T_i) \\ &= \int\frac{\left(g(t,\Lambda_0(t))+\tilde{h}(t)\right)^2}{\frac{Y(t)}{n}} d\hat{\Lambda}_{NA}(t) \xrightarrow{P} \sigma^2_{\Lambda_0}\end{aligned} \tag{8.17}$$

where $\sigma^2_{\Lambda_0} = \int \frac{\left(g(x,\Lambda_0(x))+\tilde{h}(x)\right)^2}{(1-F_0(x))(1-G_0(x))} d\Lambda_0(x)$ and $Y(t) = \sum_{i=1}^{n} I[T_i \geq t]$. The last step of (8.17) is similar to Pan and Zhou (2002) Lemma A3.

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\frac{Z_i^2}{n-i} &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n-i}\left(\frac{\delta_i\left(g(T_i,\Lambda_0(T_i))+\tilde{h}(T_i)\right)}{\frac{n-i+1}{n}}\right)^2 \\
&= \sum_{i=1}^{n}\frac{\left(g(T_i,\Lambda_0(T_i))+\tilde{h}(T_i)\right)^2}{(n-i)\left(\frac{n-i+1}{n}\right)}\frac{\delta_i}{n-i+1} \\
&= \sum_{i=1}^{n}\frac{\left(g(T_i,\Lambda_0(T_i))+\tilde{h}(T_i)\right)^2}{\frac{(Y(T_i)-1)Y(T_i)}{n}}\Delta\hat{\Lambda}_{NA}(T_i) \\
&= \int\frac{\left(g(t,\Lambda_0(t))+\tilde{h}(t)\right)}{\frac{(Y(t)-1)Y(t)}{n}}d\hat{\Lambda}_{NA}(t)\xrightarrow{P}0
\end{aligned}
\tag{8.18}
$$

The last step of (8.18) is similar to Pan and Zhou (2002) Lemma A3. $\blacksquare$

Proof of Lemma 3.2.4

Note that

$$
M_n = \max_{1\le i\le n}|Z_i| = o_p(n^{\frac{1}{2}}) \tag{8.19}
$$

By Lemma 3.2.3, λ is the solution of the following equation

$$
l(\lambda) = \frac{1}{n}\sum_{i=1}^{n-1}\frac{\delta_i Z_i}{1+\lambda Z_i} + \frac{1}{n}\delta_n Z_n - \hat{\theta} = \theta_0 \tag{8.20}
$$

where $Z_i = \frac{\delta_i(g(T_i,\Lambda_0(T_i))+\tilde{h}(T_i))}{\frac{n-i+1}{n}}$, $\quad i = 1,2,\ldots,n \quad$ and $\quad \hat{\theta} = \int\tilde{h}(t)d\Lambda_0(t)$.
Since $\delta_i Z_i = Z_i$, we have

$$
\begin{aligned}
0 &= |l(\lambda)-\theta_0| \\
&= \left|\frac{1}{n}\sum_{i=1}^{n}Z_i - \hat{\theta} - \theta_0 - \frac{1}{n}\sum_{i=1}^{n-1}\frac{\lambda Z_i^2}{1+\lambda Z_i}\right| \\
&\ge \frac{|\lambda|}{1+|\lambda|\max_{1\le i\le n}|Z_i|}\frac{1}{n}\sum_{i=1}^{n}Z_i^2 - \left|\frac{1}{n}\sum_{i=1}^{n}Z_i - \theta_0 - \hat{\theta}\right|
\end{aligned}
\tag{8.21}
$$

The second term of (8.21) is $O_p(n^{-\frac{1}{2}})$ by Lemma 8.1. Since

$$\frac{1}{n}\sum_{i=1}^{n-1} Z_i^2 = \frac{1}{n}\sum_{i=1}^{n} Z_i^2 - \frac{1}{n}Z_n^2 \tag{8.22}$$

by (8.19) we have $\frac{1}{n}Z_n^2 = o_p(1)$. Hence by Lemma 8.1.5,

$$\frac{1}{n}\sum_{i=1}^{n-1} Z_i^2 \xrightarrow{P} \int \frac{\left(g(x,\Lambda_0(x)) + \tilde{h}(x)\right)^2}{(1-F_0(x))\,(1-G_0(x))} d\Lambda_0(x) \tag{8.23}$$

this follows that

$$\frac{|\lambda|}{1+|\lambda|\max_{1\le i\le n}|Z_i|} = O_p(n^{-\frac{1}{2}}) \tag{8.24}$$

since we have $\max_{1\le i\le n}|\lambda Z_i| = o_p(1)$, then we have

$$\lambda = O_p(n^{-\frac{1}{2}}) \tag{8.25}$$

We can rewrite (3.50) as follows.

$$\begin{aligned} 0 &= l(\lambda) - \theta_0 \\ &= \frac{1}{n}\sum_{i=1}^{n} Z_i - \theta_0 - \hat{\theta} - \frac{1}{n}\sum_{i=1}^{n}\frac{\lambda Z_i^2}{1+\lambda Z_i^2} \\ &= \frac{1}{n}\sum_{i=1}^{n} Z_i - \theta_0 - \hat{\theta} - \frac{\lambda}{n}\sum_{i=1}^{n-1} Z_i^2 + \frac{1}{n}\sum_{i=1}^{n-1}\frac{\lambda^2 Z_i^3}{1+\lambda Z_i} \end{aligned} \tag{8.26}$$

The last term is bounded by

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n-1}\frac{\lambda^2 Z_i^3}{1+\lambda Z_i} &\le \frac{\lambda^2}{1-|\lambda|\max_{1\le i\le n}|Z_i|}\max_{1\le i\le n}|Z_i|\frac{1}{n}\sum_{i=1}^{n-1} Z_i^2 \\ &= O_p(n^{-1})o_p(n^{\frac{1}{2}})O_p(1) = o_p(n^{-\frac{1}{2}}) \end{aligned} \tag{8.27}$$

Then we can get an expression of λ as

$$\lambda = \frac{\frac{1}{n}\sum_{i=1}^{n} Z_i - \hat{\theta} - \theta_0}{\frac{1}{n}\sum_{i=1}^{n-1} Z_i^2} + o_p(n^{-\frac{1}{2}}) \tag{8.28}$$

By Lemma 8.1.5, Lemma 8.1, Slutsky theorem and (8.22), as $n \to \infty$

$$n\lambda^2 \xrightarrow{D} \chi^2_{(1)} \left(\int \frac{\left(g(x, \Lambda_0(x)) + \tilde{h}(x)\right)^2}{(1 - F_0(x))\,(1 - G_0(x))} d\Lambda_0(x) \right)^{-1} \tag{8.29}$$

$\blacksquare$

8.2 Additional Lemmas of Chapter 4

First of all, we prove that

$$\max_{1 \le i \le n} |Z_i| = o_p\left(\sqrt{n}\right) \tag{8.30}$$

where $Z_i = g(X_i, F_0(X_i)) + \tilde{h}(X_i) - \tilde{\theta} - \theta_0$, $\tilde{\theta} = \int \tilde{h}(t) dF_0(t)$, $\theta_0 = \int g(t, F_0(t)) dF_0(t)$, $h = \frac{\partial g(t,\Lambda)}{\partial \Lambda}$ and $\tilde{h}(t) = \int_t^{+\infty} h(s, F_0(s)) dF_0(s)$.

Under the assumption that

$$\sigma^2 = \int \left(g(t, F_0(t)) + \tilde{h}(t)\right)^2 dF_0(t) - \left(\int \left(g(t, F_0(t)) + \tilde{h}(t)\right) dF_0(t) \right)^2 < \infty \tag{8.31}$$

Since

$$E[Z]^2 = \sigma^2 < \infty \tag{8.32}$$

by Lemma 8.1.2, we have

$$\max_{1 \le i \le n} |Z_i| = o_p\left(\sqrt{n}\right) \tag{8.33}$$

Lemma 8.2.1 *Under the assumption of Lemma 4.2.3, we have, for Z_i defined in Lemma 4.2.4*

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i\right) \xrightarrow{D} N(0, \sigma^2) \tag{8.34}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} Z_i^2 \xrightarrow{P} \sigma^2 \tag{8.35}$$

where $Z_i = g(X_i, F_0(X_i)) + \tilde{h}(X_i) - \tilde{\theta} - \theta$ *and*

$$\sigma^2 = \int \left(g(t, F_0(t)) + \tilde{h}(t)\right)^2 dF_0(t) - \left(\int \left(g(t, F_0(t)) + \tilde{h}(t)\right) dF_0(t)\right)^2 \tag{8.36}$$

Proof The Central Limit Theorem and Law of Large Numbers would complete the proof of the above lemma. ▮

Proof of Lemma 4.2.5

By Lemma 4.2.4, λ is the solution of the following equation

$$l(\lambda) = \sum_{i=1}^{n} \Delta\hat{F}_n(X_i)\frac{Z_i + \tilde{\theta} + \theta_0}{1 + \lambda Z_i} = \tilde{\theta} + \theta_0 \tag{8.37}$$

where $\hat{F}_n$ is the empirical distribution function.

Therefore we have

$$\begin{aligned} 0 = \left|l(\lambda) - \tilde{\theta} - \theta_0\right| &= \left|\frac{1}{n}\sum_{i=1}^{n}\frac{Z_i}{1+\lambda Z_i} + \left(\tilde{\theta} + \theta_0\right)\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\lambda Z_i} - \left(\tilde{\theta} + \theta_0\right)\right| \\ &= \left|\frac{1}{n}\sum_{i=1}^{n} Z_i - \frac{1}{n}\sum_{i=1}^{n}\frac{\lambda Z_i^2}{1+\lambda Z_i} - \left(\tilde{\theta} + \theta_0\right)\frac{1}{n}\sum_{i=1}^{n}\frac{\lambda Z_i}{1+\lambda Z_i}\right| \\ &\geq \frac{1}{n}\sum_{i=1}^{n} Z_i - \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\lambda Z_i^2}{1+\lambda Z_i} + \left(\tilde{\theta} + \theta_0\right)\frac{1}{n}\sum_{i=1}^{n}\frac{\lambda Z_i}{1+\lambda Z_i}\right) \end{aligned} \tag{8.38}$$

By Lemma 8.2.1, we have $\frac{1}{n}\sum_{i=1}^{n} Z_i = O_p(\frac{1}{\sqrt{n}})$, then we have

$$\lambda\frac{1}{n}\sum_{i=1}^{n}\frac{Z_i^2}{1+\lambda Z_i} + \lambda\left(\tilde{\theta} + \theta_0\right)\frac{1}{n}\sum_{i=1}^{n}\frac{Z_i}{1+\lambda Z_i} = O_p(\frac{1}{\sqrt{n}}) \tag{8.39}$$

It is easy to verify that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{Z_i^2}{1+\lambda Z_i}=O_p(1),\quad \left(\tilde{\theta}+\theta_0\right)\frac{1}{n}\sum_{i=1}^{n}\frac{Z_i}{1+\lambda Z_i}=O_p(1) \tag{8.40}$$

Therefore we have

$$\lambda = O_p\left(\frac{1}{\sqrt{n}}\right) \tag{8.41}$$

On the other side, we have

$$\begin{aligned} 0 &= l(\lambda)-\tilde{\theta}-\theta_0 \\ &= \frac{1}{n}\sum_{i=1}^{n}Z_i-\lambda\frac{1}{n}\sum_{i=1}^{n}Z_i^2+\lambda^2\frac{1}{n}\sum_{i=1}^{n}\frac{Z_i^3}{1+\lambda Z_i}-\left(\tilde{\theta}+\theta_0\right)\frac{1}{n}\sum_{i=1}^{n}\frac{\lambda Z_i}{1+\lambda Z_i} \end{aligned} \tag{8.42}$$

It is easy to verify that

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}\frac{\lambda^2 Z_i^3}{1+\lambda Z_i} &\le \lambda^2\times\max_{1\le i\le n}|Z_i|\times\frac{1}{n}\sum_{i=1}^{n}Z_i^2\times\frac{1}{1+|\lambda|\max|Z_i|} \\ &\le O_P(\frac{1}{n})o_p(\sqrt{n})O_p(1)O_p(1)=o_p(\frac{1}{\sqrt{n}}) \end{aligned} \tag{8.43}$$

and

$$\left(\tilde{\theta}+\theta_0\right)\frac{1}{n}\sum_{i=1}^{n}\frac{\lambda Z_i}{1+\lambda Z_i}=O_p(\frac{1}{n}) \tag{8.44}$$

Therefore

$$\lambda=\frac{\frac{1}{n}\sum_{i=1}^{n}Z_i}{\frac{1}{n}\sum_{i=1}^{n}Z_i^2}+o_p\left(\frac{1}{\sqrt{n}}\right) \tag{8.45}$$

Multiply $\sqrt{n}$ on both sides

$$\sqrt{n}\lambda=\frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}Z_i\right)}{\frac{1}{n}\sum_{i=1}^{n}Z_i^2}+o_p(1) \tag{8.46}$$

By Lemma 8.2.1, we have

$$\sqrt{n}\lambda \xrightarrow{D} N(0,\sigma^{-2}) \tag{8.47}$$

which implies that

$$n\lambda^2 \xrightarrow{D} \chi^2_{(1)} \cdot \sigma^{-2} \tag{8.48}$$

where

$$\sigma^2 = \int \left(g(t, F_0(t)) + \tilde{h}(t)\right)^2 dF_0(t) - \left(\int \left(g(t, F_0(t)) + \tilde{h}(t)\right) dF_0(t)\right)^2 \tag{8.49}$$

The proof of Lemma 4.2.5 is finished. ■

8.3 Additional Lemmas of Chapter 5

Lemma 8.3.1 *Under the assumption of Lemma 5.2.5, we have, for* Z_{1i} *and* Z_{2j} *defined in Lemma 5.2.3,*

$$\begin{aligned}\sqrt{n}\left(\sum_{i=1}^{n} Z_{1i} - \theta_0\right) &= \sqrt{n}\int H_1(t)d\left(\hat{\Lambda}_1(t) - \Lambda_{10}(t)\right) \xrightarrow{D} N(0, \sigma_1^2), \text{ as } n \to \infty \\ \sqrt{m}\left(\sum_{j=1}^{m} Z_{2j} - \theta_0\right) &= \sqrt{m}\int H_2(s)d\left(\hat{\Lambda}_2(s) - \Lambda_{20}(s)\right) \xrightarrow{D} N(0, \sigma_2^2), \text{ as } m \to \infty\end{aligned} \tag{8.50}$$

where $\sigma_1^2 = \int \frac{H_1^2(t)}{(1-F_1(t))(1-G_1(t))} d\Lambda_{10}(t)$, $\sigma_2^2 = \int \frac{H_2^2(s)}{(1-F_2(s))(1-G_2(s))} d\Lambda_{20}(s)$ *and* $\hat{\Lambda}_1(t)$, $\hat{\Lambda}_2(s)$ *are Nelson-Aalen estimators.*

Proof First we calculate

$$\begin{aligned}\sum_{i=1}^{n} Z_{1i} - \theta_0 &= \sum_{i=1}^{n} \frac{\delta_{X_i} H_1(T_i)}{n-i+1} - \theta_0 = \sum_{i=1}^{n} H_1(T_i)\Delta\hat{\Lambda}_1(T_i) - \theta_0 \\ &= \int H_1(t)d\hat{\Lambda}_1(t) - \int H_1(t)d\Lambda_{10}(t) = \int H_1(t)d\left(\hat{\Lambda}_1(t) - \Lambda_{10}(t)\right)\end{aligned} \tag{8.51}$$

Similarly, we have

$$\sum_{j=1}^{m} Z_{2j} - \theta_0 = \int H_2(s) d\left(\hat{\Lambda}_2(s) - \Lambda_{20}(s)\right) \tag{8.52}$$

Similar arguments to Andersen et al. (1993, Chap. 4) can be used to analyze the integral. An application of the martingale central limit theorem will finish the proof.

∎

Lemma 8.3.2 *Under the assumption of Lemma 5.2.5, we have for Z_{1i} and Z_{2j} defined in Lemma 5.2.3,*

$$\begin{aligned} n\sum_{i=1}^{n} Z_{1i}^2 &\xrightarrow{P} \int \frac{H_1^2(t)}{(1-F_1(t))(1-G_1(t))} d\Lambda_{10}(t), \ \text{as } n \to \infty \\ m\sum_{j=1}^{m} Z_{2j}^2 &\xrightarrow{P} \int \frac{H_2^2(s)}{(1-F_2(s))(1-G_2(s))} d\Lambda_{20}(s), \ \text{as } m \to \infty \end{aligned} \tag{8.53}$$

Proof

$$\begin{aligned} n\sum_{i=1}^{n} Z_{1i}^2 &= n\sum_{i=1}^{n} \frac{\delta_{X_i} H_1^2(T_i)}{(n-i+1)^2} = \sum_{i=1}^{n} \frac{H_1^2(T_i)\Delta\hat{\Lambda}_1(T_i)}{\frac{n-i+1}{n}} \\ &= \int \frac{H_1^2(t)}{\frac{Y_1(t)}{n}} d\hat{\Lambda}_1(t) \xrightarrow{P} \int \frac{H_1^2(t)}{(1-F_1(t))(1-G_1(t))} d\Lambda_{10}(t) \end{aligned} \tag{8.54}$$

where $Y_1(t) = \sum_{i=1}^{n} I[T_i \geq t]$ and $\hat{\Lambda}_1$ is Nelson-Aalen estimator.

Similarly, we have

$$m\sum_{j=1}^{m} Z_{2j}^2 = \int \frac{H_2^2(s)}{\frac{Y_2(s)}{m}} d\hat{\Lambda}_2(s) \xrightarrow{P} \int \frac{H_2^2(s)}{(1-F_2(s))(1-G_2(s))} d\Lambda_{20}(s) \tag{8.55}$$

where $Y_2(s) = \sum_{j=1}^{m} I[T_j \geq s]$ and $\hat{\Lambda}_2$ is Nelson-Aalen estimator.

∎

Proof of Lemma 5.2.4

$$\begin{aligned}
0 = l(\lambda) &= \left|\sum_{i=1}^{n-1} \frac{Z_{1i}}{1+\lambda Z_{1i}} + Z_{1n} + \sum_{j=1}^{m-1} \frac{Z_{2j}}{1+\lambda Z_{2j}} + Z_{2m} - 2\theta_0\right| \\
&= \left|\sum_{i=1}^{n} Z_{1i} - \sum_{i=1}^{n-1} \frac{\lambda Z_{1i}^2}{1+\lambda Z_{1i}} + \sum_{j=1}^{m} Z_{2j} - \sum_{j=1}^{m-1} \frac{\lambda Z_{2j}^2}{1+\lambda Z_{2j}} - 2\theta_0\right| \\
&\geq \left(\frac{|\lambda|}{1+|\lambda| \max_{1\leq i\leq n} |Z_{1i}|} \sum_{i=1}^{n-1} Z_{1i}^2 + \frac{|\lambda|}{1+|\lambda| \max_{1\leq j\leq m} |Z_{2j}|} \sum_{j=1}^{m-1} Z_{2j}^2\right) \\
&- \left|\left(\sum_{i=1}^{n} Z_{1i} - \theta_0\right) + \left(\sum_{j=1}^{m} Z_{2j} - \theta_0\right)\right|
\end{aligned} \tag{8.56}$$

Since $\sum_{i=1}^{n} Z_{1i}$ and $\sum_{j=1}^{m} Z_{2j}$ are independent and we also have the following,

$$\begin{aligned}
\sqrt{n}\left(\sum_{i=1}^{n} Z_{1i} - \theta_0\right) &\xrightarrow{D} N(0, \sigma_1^2) \quad as \quad n \to \infty \\
\sqrt{m}\left(\sum_{j=1}^{m} Z_{2j} - \theta_0\right) &\xrightarrow{D} N(0, \sigma_2^2) \quad as \quad m \to \infty
\end{aligned} \tag{8.57}$$

where $\sigma_1^2 = \int \frac{H_1^2(t) d\Lambda_{10}(t)}{(1-F_1(t))(1-G_1(t))}$ and $\sigma_2^2 = \int \frac{H_2^2(s) d\Lambda_{20}(s)}{(1-F_2(s))(1-G_2(s))}$.

Assume $\frac{n}{n+m} \to \alpha$, as $\min(n, m) \to \infty$, we have

$$\sqrt{\frac{nm}{n+m}}\left(\left(\sum_{i=1}^{n} Z_{1i} - \theta_0\right) + \left(\sum_{j=1}^{m} Z_{2j} - \theta_0\right)\right) \xrightarrow{D} N(0, \sigma^2), \quad as \quad \min(n, m) \to \infty \tag{8.58}$$

where

$$\sigma^2 = (1-\alpha) \int \frac{H_1^2(t) d\Lambda_{10}(t)}{(1-F_1(t))(1-G_1(t))} + \alpha \int \frac{H_2^2(s) d\Lambda_{20}(s)}{(1-F_2(s))(1-G_2(s))} \tag{8.59}$$

Therefore

$$\sqrt{\frac{nm}{n+m}}\left|\left(\sum_{i=1}^{n} Z_{1i} - \theta_0\right) + \left(\sum_{j=1}^{m} Z_{2j} - \theta_0\right)\right| = O_p(1) \tag{8.60}$$

which means

$$\left(\frac{|\lambda|}{1+|\lambda|\max_{1\leq i\leq n}|Z_{1i}|}\sum_{i=1}^{n-1}Z_{1i}^2+\frac{|\lambda|}{1+|\lambda|\max_{1\leq j\leq m}|Z_{2j}|}\sum_{j=1}^{m-1}Z_{2j}^2\right)$$
$$=\left(\frac{\frac{1}{n}|\lambda|}{1+\frac{1}{n}|\lambda|\max_{1\leq i\leq n}(n|Z_{1i}|)}n\sum_{i=1}^{n-1}Z_{1i}^2+\frac{\frac{1}{m}|\lambda|}{1+\frac{1}{m}|\lambda|\max_{1\leq j\leq m}(m|Z_{2j}|)}m\sum_{j=1}^{m-1}Z_{2j}^2\right)$$
$$=O_p\left(\sqrt{\frac{n+m}{nm}}\right) \tag{8.61}$$

It can be shown that $\max_{1\leq i\leq n}(n|Z_{1i}|)=o_p(\sqrt{n})$ and $\max_{1\leq j\leq m}(m|Z_{2j}|)=o_p(\sqrt{m})$ and by the Lemma 8.3.2, we have $n\sum_{i=1}^{n-1}Z_{1i}^2=O_p(1)$ and $m\sum_{j=1}^{m-1}Z_{2j}^2=O_p(1)$. Then we can conclude that

$$\frac{1}{n}|\lambda|+\frac{1}{m}|\lambda|=O_p\left(\sqrt{\frac{n+m}{nm}}\right) \tag{8.62}$$

which means

$$|\lambda|=O_p\left(\left(\frac{n+m}{nm}\right)^{-\frac{1}{2}}\right) \tag{8.63}$$

Now we have

$$\begin{aligned}0=l(\lambda)=&\left(\sum_{i=1}^{n}Z_{1i}-\theta_0\right)-\lambda\sum_{i=1}^{n-1}Z_{1i}^2+\sum_{i=1}^{n-1}\frac{\lambda^2Z_{1i}^3}{1+\lambda Z_{1i}}\\&+\left(\sum_{j=1}^{m}Z_{2j}-\theta_0\right)-\lambda\sum_{j=1}^{m-1}Z_{2j}^2+\sum_{j=1}^{m-1}\frac{\lambda^2Z_{1j}^3}{1+\lambda Z_{2j}}\end{aligned} \tag{8.64}$$

We also have

$$\sum_{i=1}^{n-1}\frac{\lambda^2Z_{1i}^3}{1+\lambda Z_{1i}}\leq\sum_{i=1}^{n-1}\frac{\lambda^2Z_{1i}^3}{1-|\lambda|\max|Z_{1i}|}\leq\left(\frac{1}{n^2}\lambda^2\times\max|nZ_{1i}|\times n\sum_{i=1}^{n-1}Z_{1i}^2\right)O_p(1)$$
$$=o_p(\frac{1}{\sqrt{n}})=o_p\left(\sqrt{\frac{n+m}{nm}}\right) \tag{8.65}$$

Similarly,

$$\sum_{j=1}^{m-1} \frac{\lambda^2 Z_{2j}^3}{1+\lambda Z_{2j}} \leq o_p\left(\frac{1}{\sqrt{m}}\right) = o_p\left(\sqrt{\frac{n+m}{nm}}\right) \tag{8.66}$$

So we have an expression of λ,

$$\lambda = \frac{\left(\sum_{i=1}^{n} Z_{1i} - \theta_0\right) + \left(\sum_{j=1}^{m} Z_{2j} - \theta_0\right)}{\sum_{i=1}^{n-1} Z_{1i}^2 + \sum_{j=1}^{m-1} Z_{2j}^2} + o_p\left(\sqrt{\frac{nm}{n+m}}\right) \tag{8.67}$$

Multiplying $\sqrt{\frac{n+m}{nm}}$ on each side gives us

$$\sqrt{\frac{n+m}{nm}}\lambda = \frac{\sqrt{\frac{nm}{n+m}}\left(\sum_{i=1}^{n} Z_{1i} - \theta_0 + \sum_{j=1}^{m} Z_{2j} - \theta_0\right)}{\frac{nm}{n+m}\left(\sum_{i=1}^{n-1} Z_{1i}^2 + \sum_{j=1}^{m-1} Z_{2j}^2\right)} + o_p(1) \tag{8.68}$$

Since

$$\sqrt{\frac{nm}{n+m}}\left(\sum_{i=1}^{n} Z_{1i} - \theta_0 + \sum_{j=1}^{m} Z_{2j} - \theta_0\right) \xrightarrow{D} N(0, \sigma^2) \tag{8.69}$$

where $\sigma^2 = (1-\alpha)\int \frac{H_1^2(t)d\Lambda_{10}(t)}{(1-F_1(t))(1-G_1(t))} + \alpha \int \frac{H_2^2(s)d\Lambda_{20}(s)}{(1-F_2(s))(1-G_2(s))}$ and

$$\begin{aligned}\frac{nm}{n+m}\left(\sum_{i=1}^{n-1} Z_{1i}^2 + \sum_{j=1}^{m-1} Z_{2j}^2\right) \xrightarrow{\mathscr{D}} (1-\alpha)\int \frac{H_1^2(t)d\Lambda_{10}(t)}{(1-F_1(t))(1-G_1(t))} \\ + \alpha \int \frac{H_2^2(s)d\Lambda_{20}(s)}{(1-F_2(s))(1-G_2(s))} = \sigma^2\end{aligned} \tag{8.70}$$

By the Slutsky theorem,

$$\frac{n+m}{nm}\lambda^2 \xrightarrow{D} \chi_1^2\sigma^{-2}, \quad as \quad \min(n,m) \to \infty \tag{8.71}$$

where $\sigma^2 = (1-\alpha)\int \frac{H_1^2(t)d\Lambda_{10}(t)}{(1-F_1(t))(1-G_1(t))} + \alpha \int \frac{H_2^2(s)d\Lambda_{20}(s)}{(1-F_2(s))(1-G_2(s))}$ ∎

Lemma 8.3.3 *Under the assumption of Lemma 5.3.4, we have, for* Z_{1i} *and* Z_{2j}

defined in Lemma 5.3.2,

$$\begin{aligned}
\sqrt{n}\left(\sum_{i=1}^{n} Z_{1i} - \tilde{\theta}_1 - \theta_0\right) &\xrightarrow{D} N(0,\sigma_1^2) \;\; as \;\; n \to \infty \\
\sqrt{m}\left(\sum_{j=1}^{m} Z_{2j} - \tilde{\theta}_2 - \theta_0\right) &\xrightarrow{D} N(0,\sigma_2^2) \;\; as \;\; m \to \infty
\end{aligned} \tag{8.72}$$

where $\sigma_1^2 = \int \frac{H_1^{*2}(t)d\Lambda_{10}(t)}{(1-F_1(t))(1-G_1(t))}$ *and* $\sigma_2^2 = \int \frac{H_2^{*2}(s)d\Lambda_{20}(s)}{(1-F_2(s))(1-G_2(s))}$.

Proof

$$\begin{aligned}
&\sqrt{n}\left(\sum_{i=1}^{n} Z_{1i} - \tilde{\theta}_1 - \theta_0\right) \\
&= \sqrt{n}\left(\sum_{i=1}^{n} \frac{\delta_{X_i} H_1^*(T_i)}{n-i+1} - \tilde{\theta}_1 - \theta_0\right) = \sqrt{n}\left(\int H_1^*(t)d\hat{\Lambda}_1(t) - \tilde{\theta}_1 - \theta_0\right) \\
&= \sqrt{n}\left(\int\int \left(H + \tilde{H}_1\right) d\Lambda_{20} d\hat{\Lambda}_1 - \int\int \tilde{H}_1 d\Lambda_{10} d\Lambda_{20} - \int\int H d\Lambda_{10} d\Lambda_{20}\right) \\
&= \sqrt{n}\left(\int\int \left(H + \tilde{H}_1\right) d\Lambda_{20} d\left(\hat{\Lambda}_1 - \Lambda_{10}\right)\right) \\
&= \sqrt{n}\left(\int \tilde{H}_1^*(t) d\left(\hat{\Lambda}_1(t) - \Lambda_{10}(t)\right)\right) \xrightarrow{D} N(0,\sigma_1^2)
\end{aligned} \tag{8.73}$$

where $\sigma_1^2 = \int \frac{H_1^{*2}(t)d\Lambda_{10}(t)}{(1-F_1(t))(1-G_1(t))}$.

Similarly,

$$\sqrt{m}\left(\sum_{j=1}^{m} Z_{2j} - \tilde{\theta}_2 - \theta_0\right) = \sqrt{m}\int H_2^*(s)d\left(\hat{\Lambda}_2(s) - \Lambda_{20}(s)\right) \xrightarrow{D} N(0,\sigma_2^2) \tag{8.74}$$

where $\sigma_2^2 = \int \frac{H_2^{*2}(s)d\Lambda_{20}(s)}{(1-F_2(s))(1-G_2(s))}$.

∎

Lemma 8.3.4 *Under the assumption of Lemma 5.3.4, we have, for* Z_{1i} *and* Z_{2j}

defined in Lemma 5.3.2,

$$n\sum_{i=1}^{n} Z_{1i}^2 \xrightarrow{P} \int \frac{H_1^{*2}(t)}{(1-F_1(t))\,(1-G_1(t))} d\Lambda_{10}(t)$$
$$m\sum_{j=1}^{m} Z_{2j}^2 \xrightarrow{P} \int \frac{H_2^{*2}(s)}{(1-F_2(s))\,(1-G_2(s))} d\Lambda_{20}(s) \tag{8.75}$$

Proof Assume $\frac{n}{n+m} \to \alpha,\ as\ \min(n,m) \to \infty$

$$n\sum_{i=1}^{n} Z_{1i}^2 = n\sum_{i=1}^{n} \frac{\delta_{X_i} H_1^{*2}(T_i)}{(n-i+1)^2}$$
$$= \int \frac{H_1^{*2}(t)}{\frac{Y_1(t)}{n}} d\hat{\Lambda}_1(t) \xrightarrow{P} \int \frac{H_1^{*2}(t)}{(1-F_1(t))\,(1-G_1(t))} d\Lambda_{10}(t) \tag{8.76}$$

where $Y_1(t) = \sum_{i=1}^{n} I[T_i \geq t]$.

Similarly,

$$m\sum_{j=1}^{m} Z_{2j}^2 = \int \frac{H_2^{*2}(s)}{\frac{Y_2(s)}{n}} d\hat{\Lambda}_2(s) \xrightarrow{P} \int \frac{H_2^{*2}(s)}{(1-F_2(s))\,(1-G_2(s))} d\Lambda_{20}(s) \tag{8.77}$$

where $Y_2(s) = \sum_{j=1}^{m} I[U_j \geq s]$.

∎

Proof of Lemma 5.3.3

$$0 = l(\lambda) = \sum_{i=1}^{n-1} \frac{Z_{1i}}{1+\lambda Z_{1i}} + Z_{1n} + \sum_{j=1}^{m-1} \frac{Z_{2j}}{1+\lambda Z_{2j}} + Z_{2m} - \tilde{\theta}_1 - \tilde{\theta}_2 - 2\theta_0$$
$$= \left(\sum_{i=1}^{n} Z_{1i} - \tilde{\theta}_1 - \theta_0\right) + \left(\sum_{j=1}^{m} Z_{2j} - \tilde{\theta}_2 - \theta_0\right) - \lambda\sum_{i=1}^{n-1} Z_{1i}^2 - \lambda\sum_{j=1}^{m-1} Z_{2j}^2 \tag{8.78}$$
$$+ \sum_{i=1}^{n-1} \frac{\lambda^2 Z_{1i}^3}{1+\lambda Z_{1i}} + \sum_{j=1}^{m-1} \frac{\lambda^2 Z_{2j}^3}{1+\lambda Z_{2j}}$$

It can be shown that

$$\begin{aligned}\sum_{i=1}^{n-1} \frac{\lambda^2 Z_{1i}^3}{1+\lambda Z_{1i}} &= o_p\left(\sqrt{\frac{n+m}{nm}}\right)\\ \sum_{j=1}^{m-1} \frac{\lambda^2 Z_{2j}^3}{1+\lambda Z_{2j}} &= o_p\left(\sqrt{\frac{n+m}{nm}}\right)\end{aligned} \tag{8.79}$$

Since

$$\begin{aligned}\frac{nm}{n+m}\left(\sum_{i=1}^{n-1} Z_{1i}^2 + \sum_{j=1}^{m-1} Z_{2j}^2\right) \xrightarrow{P} (1-\alpha)\int \frac{H_1^{*2}(t)}{(1-F_1(t))\,(1-G_1(t))} d\Lambda_{10}\\ + \alpha \int \frac{H_2^{*2}(s)}{(1-F_2(s))\,(1-G_2(s))} d\Lambda_{20}\end{aligned} \tag{8.80}$$

we have

$$\left|\sum_{i=1}^{n-1} Z_{1i}^2 + \sum_{j=1}^{m-1} Z_{2j}^2\right| = O_p\left(\frac{n+m}{nm}\right) \tag{8.81}$$

Now we have an expression of λ as follows,

$$\lambda = \frac{\left(\sum_{i=1}^n Z_{1i} - \tilde{\theta}_1 - \theta_0\right) + \left(\sum_{j=1}^m Z_{2j} - \tilde{\theta}_2 - \theta_0\right)}{\sum_{i=1}^{n-1} Z_{1i}^2 + \sum_{j=1}^{m-1} Z_{2j}^2} + o_p\left(\sqrt{\frac{nm}{n+m}}\right) \tag{8.82}$$

Since $\sum_{i=1}^n Z_{1i}$ and $\sum_{j=1}^m Z_{2j}$ are independent, by the Lemma 8.3.3, we have

$$\begin{aligned}&\sqrt{\frac{nm}{n+m}}\left(\sum_{i=1}^n Z_{1i} - \tilde{\theta}_1 - \theta_0 + \sum_{j=1}^m Z_{2j} - \tilde{\theta}_2 - \theta_0\right) \xrightarrow{D} N(0,\sigma^2),\\ &as \quad \min(n,m) \ \to \infty\end{aligned} \tag{8.83}$$

where $\sigma^2 = (1-\alpha)\int \frac{H_1^{*2}(t)}{(1-F_1(t))(1-G_1(t))} d\Lambda_{10} + \alpha \int \frac{H_2^{*2}(s)}{(1-F_2(s))(1-G_2(s))} d\Lambda_{20}$.

And we also have

$$\frac{nm}{n+m}\left(\sum_{i=1}^{n-1} Z_{1i}^2 + \sum_{j=1}^{m-1} Z_{2j}^2\right) \xrightarrow{P} (1-\alpha)\int \frac{H_1^{*2}(t)}{(1-F_1(t))(1-G_1(t))} d\Lambda_{10} + \alpha \int \frac{H_2^{*2}(s)}{(1-F_2(s))(1-G_2(s))} d\Lambda_{20} \tag{8.84}$$

By the Slutsky theorem, we have

$$\frac{n+m}{nm}\lambda^2 \xrightarrow{D} \chi_1^2 \sigma^{-2}, \quad as \quad \min(n,m) \quad \to \infty \tag{8.85}$$

∎

8.4 R code

```
### Libraries needed ###

library(emplik)

library(survival)

library(KMsurv)

library(actuar)


### function with g(t) independent of \Lambda(t);

### modified from emplikH1.test

emplikh1.test<-function (x, d, y = -Inf, theta, fun,

tola = .Machine$double.eps^0.5)

{

  n <- length(x)

  if (n <= 2)

    stop("Need more observations")
```

```
if (length(d) != n)
  stop("length of x and d must agree")
if (any((d != 0) & (d != 1)))
  stop("d must be 0/1's for censor/not-censor")
if (!is.numeric(x))
  stop("x must be numeric values --- observed times")
newdata <- Wdataclean2(x, d)
temp <- DnR(newdata$value, newdata$dd, newdata$weight, y = y)
time <- temp$times
risk <- temp$n.risk
jump <- (temp$n.event)/risk
funtime <- fun(time)
funh <- sqrt(n) * funtime/risk
funtimeTjump <- funtime * jump
if (jump[length(jump)] >= 1)
  funh[length(jump)] <- 0
inthaz <- function(x, ftj, fh, thet) {
  sum(ftj/(1 + x * fh)) - thet
}
diff <- inthaz(0, funtimeTjump, funh, theta)
if (diff == 0) {
  lam <- 0
}
else {
  step <- 0.01/sqrt(n)
  mini <- 0
  maxi <- 0
```

```
    if (diff > 0) {
      maxi <- step
      while (inthaz(maxi, funtimeTjump, funh, theta) >
                0 && maxi < 1000) maxi <- maxi +
        step
    }
    else {
      mini <- -step
      while (inthaz(mini, funtimeTjump, funh, theta) <
                0 && mini > -1000) mini <- mini -
        step
    }
    if (inthaz(mini, funtimeTjump, funh, theta) * inthaz(maxi,
       funtimeTjump, funh, theta) > 0)
      stop("given theta is too far away from theta0")
    temp2 <- uniroot(inthaz, c(mini, maxi), tol = tola,
    ftj = funtimeTjump, fh = funh, thet = theta)
    lam <- temp2$root
  }
  onepluslamh <- 1 + lam * funh
  weights <- jump/onepluslamh
  loglik <- 2 * (sum(log(onepluslamh)) - sum((onepluslamh -
                                                  1)/onepluslamh))
  list(‘-2LLR‘ = loglik, lambda = lam/sqrt(n), times = time,
       wts = weights, nits = temp2$nf, message = temp2$message)
}
```

```
### Function when \Lambda(t) is fixed in g(t,\Lambda(t)) ###
emplikGH1.test<-function(t,d,fun1,theta1) {
  newdata <- Wdataclean2(t, d)
  time <- newdata$value
  cen<-newdata$dd
  old_lambda<-0
  old_llr<-0
  old_jump <- cen/(length(time):1)
  old_jump<-old_jump[old_jump!=0]
  gg<-function(x,jp=old_jump){
    return(fun1(x,jp))
  }
  run0<-emplikh1.test(t,d,fun=gg,theta=theta1)
  new_lambda<-run0$lambda
  new_jump<-run0$wts
  new_llr<-run0$"-2LLR"

  while(sqrt(sum((abs(old_jump-new_jump))^2)) >=1e-12
  || abs(old_lambda-new_lambda)>=1e-12
  || abs(old_llr-new_llr)>1e-12) {
    old_llr<-new_llr
    old_lambda<-new_lambda
    old_jump<-new_jump
    gg<-function(x,jp=old_jump){
      return(fun1(x,jp))
    }
    run<-emplikh1.test(t,d,theta=theta1,fun=gg)
```

```
    new_lambda<-run$lambda
    new_jump<-run$wts
    new_llr<-run$"-2LLR"
    print(new_llr)
  }
  list(`-2LLR` = run$'-2LLR', lambda = new_lambda,
  times = time,jump=new_jump)
}

### Function to generate QQ plot and track time ###
myqqplot1<-function(n,m,fun,theta,title) {
  time1<-proc.time()
  elr<-rep(NA,m)
  for (j in 1:m) {
    x<-rexp(n,rate=1)
    c<-rexp(n,rate=0.35)
    t<-rep(NA,n)
    d<-rep(NA,n)
    for (i in 1:n) {
      t[i]<-min(x[i],c[i])
      if (x[i]<c[i]) d[i]<-1
      else d[i]<-0
    }
    run<-emplikGH1.test(t,d,fun,theta)
    elr[j]<-run$'-2LLR'
  }
  y<-qchisq(seq(1/m,1,1/m)-1/(2*m),df=1)
```

```
  plot(sort(elr),y,xlab="Sample quantile",
  ylab="Chi-square theoretical quantile"
  ,main=title)
  abline(a=0,b=1)
  proc.time()-time1
}

### Function to calculate coverage probability
### of empirical likelihood method
coverage<-function(n,m,fun,theta) {
  count<-rep(0,m)
  for (j in 1:m) {
    x<-rexp(n,rate=1)
    c<-rexp(n,rate=0.35)
    t<-rep(NA,n)
    d<-rep(NA,n)
    for (i in 1:n) {
      t[i]<-min(x[i],c[i])
      if (x[i]<c[i]) d[i]<-1
      else d[i]<-0
    }
    run<-emplikGH1.test(t,d,fun,theta)
    if (run$'-2LLR'<=qchisq(0.95,df=1))
      count[j]<-1
  }
  return (mean(count))
}
```

```
### The function to generate Wald confidence
### interval with different transformations
waldci<-function(x,d,a,alpha,type) {
  my.fit<-survfit(Surv(x,d)~1)
  km<-stepfun(my.fit$time,c(1,my.fit$surv))
  ni<-my.fit$n.risk
  di<-my.fit$n.event
  jump<-di/(ni*(ni-di))
  greenwood<-stepfun(my.fit$time,c(0,cumsum(jump)))
  kmest<-km(a)
  sigma<-sqrt(greenwood(a))
  z<-qnorm(c(alpha/2,1-alpha/2))
  asest<-asin(sqrt(km(a)))
  temp<-asest+0.5*z*sigma*(km(a)/(1-km(a)))^(0.5)
  aslow<-(sin(max(0,temp[1])))^2
  asup<-(sin(min(pi/2,temp[2])))^2
  if (type=='plain')
    return (kmest+z*kmest*sigma)
  if (type=='log')
    return (kmest*exp(z*sigma/kmest))
  if (type=='log(-log)')
    return (kmest^(exp(z*sigma/(log(kmest)))))
  if (type=='arcsin')
    return (c(aslow,asup))
}
```

```
### The function to generate empirical
### likelihood confidence interval
elrci<-function(t,d,fun,alpha,lower,upper,by,step) {
  theta<-seq(lower,upper,by)
  ELR<-rep(NA,length(theta))
  for (i in 1:length(theta)) {
    ELR[i]<-emplikGH1.test(t,d,fun,theta[i])$'-2LLR'
  }
  Lower<-min(theta[ELR<=qchisq(1-alpha,1)])
  Upper<-max(theta[ELR<=qchisq(1-alpha,1)])
  l<-lower
  u<-upper
  while (Lower==theta[1] || Upper==theta[length(theta)]) {
    if (Lower==theta[1] && Upper!=theta[length(theta)]) {
      l<-l-step
      theta2<-seq(l+by,l+step-by,by)
      ELR2<-rep(NA,length(theta2))
      for (i in 1:length(theta2)) {
        ELR2[i]<-emplikGH1.test(t,d,fun,theta2[i])$'-2LLR'
      }
      theta<-c(theta2,theta)
      ELR<-c(ELR2,ELR)
    }
    else if (Lower!=theta[1] && Upper==theta[length(theta)]) {
      u<-u+step
      theta2<-seq(u-step+by,u-by,by)
      ELR2<-rep(NA,length(theta2))
```

```
    for (i in 1:length(theta2)) {
      ELR2[i]<-emplikGH1.test(t,d,fun,theta2[i])$'-2LLR'
    }
    theta<-c(theta,theta2)
    ELR<-c(ELR,ELR2)
  }
  else if (Lower==theta[1] && Upper==theta[length(theta)]) {
    l<-l-step
    theta2<-seq(l+by,l+step-by,by)
    ELR2<-rep(NA,length(theta2))
    for (i in 1:length(theta2)) {
      ELR2[i]<-emplikGH1.test(t,d,fun,theta2[i])$'-2LLR'
    }
    theta<-c(theta2,theta)
    ELR<-c(ELR2,ELR)
    u<-u+step
    theta2<-seq(u-step+by,u-by,by)
    ELR2<-rep(NA,length(theta2))
    for (i in 1:length(theta2)) {
      ELR2[i]<-emplikGH1.test(t,d,fun,theta2[i])$'-2LLR'
    }
    theta<-c(theta,theta2)
    ELR<-c(ELR,ELR2)
  }
  Lower<-min(theta[ELR<=qchisq(1-alpha,1)])
  Upper<-max(theta[ELR<=qchisq(1-alpha,1)])
}
```

```
  plot(theta,ELR,cex=0.5,xlab="Theta",ylab="-2logELR",
       main="Empirical Likelihood Ratio Confidence Interval")
  abline(h=3.84)
  abline(v=Lower)
  abline(v=Upper)
  return (c(Lower,Upper))
}

### The function to compare the coverage
### probability using different methods
CP<-function(n,m,a,fun,theta,alpha,lower,upper,by,step) {
  mu<-exp(-a)
  count1<-0
  count2<-0
  count3<-0
  count4<-0
  count5<-0
  for (j in 1:m) {
    x<-rexp(n,rate=1)
    c<-rexp(n,rate=0.35)
    t<-rep(NA,n)
    d<-rep(NA,n)
    for (i in 1:n) {
      t[i]<-min(x[i],c[i])
      if (x[i]<c[i]) d[i]<-1
      else d[i]<-0
    }
```

```
    elr<-emplikGH1.test(t,d,fun,1-exp(-a))$'-2LLR'
    ci2<-waldci(t,d,a,alpha,'plain')
    ci3<-waldci(t,d,a,alpha,'log')
    ci4<-waldci(t,d,a,alpha,'log(-log)')
    ci5<-waldci(t,d,a,alpha,'arcsin')
    if (elr <= qchisq(1-alpha,df=1))
      count1<-count1+1
    if (mu <= ci2[2] && mu >= ci2[1])
      count2<-count2+1
    if (mu <= ci3[2] && mu >= ci3[1])
      count3<-count3+1
    if (mu <= ci4[2] && mu >= ci4[1])
      count4<-count4+1
    if (mu <= ci5[2] && mu >= ci5[1])
      count5<-count5+1
  }
  return (list(cp_ELR=count1/m,cp_Normal=count2/m,cp_Log=count3/m,
               cp_Loglog=count4/m,cp_Arcsin=count5/m))
}

### The function to compare the coverage probability ###
### and average length using different methods ###
CPandAVL<-function(n,m,a,fun,theta,alpha,lower,upper,by,step) {
  mu<-exp(-a)
  count1<-0
  count2<-0
  count3<-0
```

```
count4<-0
count5<-0
AVL1<-rep(NA,m)
AVL2<-rep(NA,m)
AVL3<-rep(NA,m)
AVL4<-rep(NA,m)
AVL5<-rep(NA,m)
time1<-proc.time()
for (j in 1:m) {
  tm1<-proc.time()
  x<-rexp(n,rate=1)
  c<-rexp(n,rate=0.35)
  t<-rep(NA,n)
  d<-rep(NA,n)
  for (i in 1:n) {
    t[i]<-min(x[i],c[i])
    if (x[i]<c[i]) d[i]<-1
    else d[i]<-0
  }
  elr<-emplikGH1.test(t,d,fun,1-exp(-a))$'-2LLR'
  ci1<-elrci(t,d,fun,alpha,lower,upper,by,step)
  ci2<-waldci(t,d,a,alpha,'plain')
  ci3<-waldci(t,d,a,alpha,'log')
  ci4<-waldci(t,d,a,alpha,'log(-log)')
  ci5<-waldci(t,d,a,alpha,'arcsin')
  if (elr <= qchisq(1-alpha,df=1))
    count1<-count1+1
```

```
    if (mu <= ci2[2] && mu >= ci2[1])
      count2<-count2+1
    if (mu <= ci3[2] && mu >= ci3[1])
      count3<-count3+1
    if (mu <= ci4[2] && mu >= ci4[1])
      count4<-count4+1
    if (mu <= ci5[2] && mu >= ci5[1])
      count5<-count5+1
    AVL1[j]<-ci1[2]-ci1[1]
    AVL2[j]<-ci2[2]-ci2[1]
    AVL3[j]<-ci3[2]-ci3[1]
    AVL4[j]<-ci4[2]-ci4[1]
    AVL5[j]<-ci5[2]-ci5[1]
    tm2<-proc.time()
    print(c(j,tm2-tm1))
  }
  time2<-proc.time()
  return (list(Time=time2-time1,cp_ELR=count1/m,
  cp_Normal=count2/m,cp_Log=count3/m,
  cp_Loglog=count4/m,cp_Arcsin=count5/m,AVL_ELR=mean(AVL1),
  AVL_Normal=mean(AVL2),AVL_Log=mean(AVL3),AVL_Loglog=mean(AVL4),
  AVL_Arcsin=mean(AVL5),sd_ELR=sd(AVL1),sd_Normal=sd(AVL2),
  sd_Log=sd(AVL3),sd_Loglog=sd(AVL4),sd_arcsin=sd(AVL5)) )
}

### The functions of g(t,\Lambda(t)) ###
g1<-function(x,jp,a=0.5) {
```

```
  return (as.numeric((x <= a))*exp(-cumsum(jp)))
}
g2<-function(x,jp){
  return (exp(-(x+cumsum(jp))))
}


### Gini’s absolute mean difference-right continuous version ###
ginimdf<-function(x,jp){
  return ((2*cumsum(jp)-1)*x)
}
### Gini’s absolute mean difference-left continuous version ###
ginimdf2<-function(x,jp) {
  return ((2*c(0,cumsum(jp)[1:(length(jp)-1)])-1)*x)
}
### Gini’s absolute mean difference-middle point version ###
ginimdf3<-function(x,jp) {
  temp1<-cumsum(jp)
  temp2<-c(0,temp1[1:(length(jp)-1)])
  temp3<-(temp1+temp2)/2
  return ((2*temp3-1)*x)
}

### The function to compare the bias of three versions of ###
### Gini’s absolute mean difference ###
Bias<-function(n,m,mu,dist) {
  jump0<-rep(1/n,n)
```

```
  bias1<-rep(NA,m)
  bias2<-rep(NA,m)
  bias3<-rep(NA,m)
  for (j in 1:m) {
    if (dist=="exp")
       x<-rexp(n,rate=1)
    if (dist=="chisq")
       x<-rchisq(n,df=1)
    if (dist=="lnorm")
       x<-rlnorm(n)
    y<-sort(x)
    bias1[j]<-sum(ginimdf(y,jump0)*jump0)-mu
    bias2[j]<-sum(ginimdf2(y,jump0)*jump0)-mu
    bias3[j]<-sum(ginimdf3(y,jump0)*jump0)-mu
  }
  list(BiasR=mean(bias1),BiasL=mean(bias2),BiasM=mean(bias3))
}


#### The function of our iterative algorithm:
###  modified from el.cen.EM
el.cen.EM.G<-function(x,d,fun,mu,err) {
  n<-length(x)
  old_jump<-rep(1/n,n)
  old_loglik<-0
  fit<-el.cen.EM(x,d,fun,mu,jp=old_jump)
  new_loglik<-fit$loglik
```

```
  new_jump<-fit$prob
  while (abs(old_loglik-new_loglik)>=err
  || mean(abs(old_jump-new_jump)) >=err) {
    old_loglik<-new_loglik
    old_jump<-new_jump
    fit<-el.cen.EM(x,d,fun,mu,jp=old_jump)
    new_jump<-fit$prob
    new_loglik<-fit$loglik
  }
  bias<-sum(fun(fit$times,fit$prob)*fit$prob)-mu
  list(bias=bias,loglik=fit$loglik,times=fit$times,
  prob=fit$prob,lam=fit$lam,
  '-2LLR'=fit$'-2LLR',Pval=fit$Pval)
}

#### The function of Qin et al.'s mehtod (2011) ###
el.cen.EM.G2<-function(x,d,fun1,fun2,mu) {
  n<-length(x)
  jump0<-rep(1/n,n)
  fit<-el.cen.EM(x,d,fun1,mu,jp=jump0)
  tm<-fit$times
  jp<-fit$prob
  u1<-2*fun2(tm,jump0)-tm
  var3<-var(u1)
  u2<-fun1(tm,jump0)
  var2<-var(u2)
  khat<-var2/var3
```

```
  adjllr<-khat*fit$'-2LLR'
  list(adjllr=adjllr,loglik=fit$loglik,times=fit$times,
  prob=fit$prob,lam=fit$lam,
  '-2LLR'=fit$'-2LLR',Pval=fit$Pval)
}


### The functions of Peng's mehtod (2011) ###
sollam<-function(lam,x) {
  return (mean(x/(1+lam*x)))
}

llr<-function(theta,y,z,step) {
  temp<-y-z-theta
  num0<-sollam(0,temp)
  min<-0
  max<-0
  if (abs(sollam(step,temp)) < abs(sollam(-step,temp))) {
    max<-step
    while(num0*sollam(max,temp)>=0) {
      max<-max+step
    }
  }
  else {
    min<-(-step)
    while(num0*sollam(min,temp)>=0) {
      min<-min-step
```

```
    }
  }
  lambda<-uniroot(sollam,c(min,max),check.conv=T,x=temp)$root
  return (2*sum(log(1+lambda*temp)))
}

### The function to generate the QQ plot using our method ###
myqqplot2<-function(n,m,fun,mu,err,title) {
  elr<-rep(NA,m)
  d<-rep(1,n)
  for (j in 1:m) {
    x<-rchisq(n,df=1)
    fit<-el.cen.EM.G(x,d,fun,mu,err)
    elr[j]<-fit$'-2LLR'
    print(j)
  }
  yy<-qchisq(seq(1/m,1,1/m)-1/(2*m),df=1)
  plot(sort(elr),yy,xlab="Sample quantile",
  ylab="Chi-square theoretical quantile"
  ,main=title)
  abline(a=0,b=1)
}

### The function to compare the coverage
### probabilities of three methods
CP_Comp<-function(n,m,mu,fun1,fun2,err,sp,dist,cali=FALSE) {
  mid<-n/2
```

```
d<-rep(1,n)
elr1<-rep(NA,m)
elr2<-rep(NA,m)
elr3<-rep(NA,m)
for (j in 1:m) {
  if (dist=="exp")
     x<-rexp(n,rate=1)
  if (dist=="chisq")
     x<-rchisq(n,df=1)
  if (dist=="lnorm")
     x<-rlnorm(n)
  y<-(x[1:mid]+x[(mid+1):n])/2
  z<-rep(NA,mid)
  for (i in 1:mid) {
    z[i]<-min(x[i],x[mid+i])
  }
  fit1<-el.cen.EM.G(x,d,fun1,mu,err)
  elr1[j]<-fit1$'-2LLR'
  fit2<-el.cen.EM.G2(x,d,fun1,fun2,mu)
  elr2[j]<-fit2$adjllr
  elr3[j]<-llr(mu,y,z,sp)
  if (cali)
    quan<-qf(0.95,1,n-2)
  else
    quan<-qchisq(0.95,1)
}
list(CP1=mean(elr1<=quan),CP2=mean(elr2<=qchisq(0.95,1)),
```

```
  CP3=mean(elr3<=qchisq(0.95,1)))
}

### The functions to estimete Gini index
### and Gini's absolute mean difference
Giniest<-function(x){
  n<-length(x)
  tm<-sort(x)
  jump0<-rep(1/n,n)
  mu<-mean(x)
  return ((2*mean((2*cumsum(jump0)-1)*tm))/(2*mu))
}
GMDest<-function(x){
  n<-length(x)
  tm<-sort(x)
  jump0<-rep(1/n,n)
  return (2*mean((2*cumsum(jump0)-1)*tm))
}

### The function used in Qin et al's (2011) method ###
h1<-function(x,jp) {
  n<-length(jp)
  temp1<-x*cumsum(jp)
  temp2<-rev(cumsum(rev(x[1:n])))/n
  return (temp1+temp2)
}
```

```
### The function used in Peng's method ###
f<-function(x,Mu) {
  n<-length(x)
  mid<-n/2
  y<-(x[1:mid]+x[(mid+1):n])/2
  z<-rep(NA,mid)
  for (i in 1:mid) {
    z[i]<-min(x[i],x[mid+i])
  }
  return (y-z-y*Mu)
}

### The function to generate EM
### confidence interval using our method
findELci<-function(x,fun,err,alpha,lower,upper,by,step,cali=FALSE) {
  n<-length(x)
  d<-rep(1,n)
  theta<-seq(lower,upper,by)
  ELR<-rep(NA,length(theta))
  if (cali)
    quan<-qf(1-alpha,1,n-2)
  else
    quan<-qchisq(1-alpha,1)

  for (i in 1:length(theta)) {
    ELR[i]<-el.cen.EM.G(x,d,fun,theta[i],err)$'-2LLR'
  }
```

```
Lower<-min(theta[ELR<=quan])
Upper<-max(theta[ELR<=quan])
l<-lower
u<-upper
while (Lower==theta[1] || Upper==theta[length(theta)]) {
  if (Lower==theta[1] && Upper!=theta[length(theta)]) {
    l<-l-step
    theta2<-seq(l+by,l+step-by,by)
    ELR2<-rep(NA,length(theta2))
    for (i in 1:length(theta2)) {
      ELR2[i]<-el.cen.EM.G(x,d,fun,theta2[i],err)$'-2LLR'
    }
    theta<-c(theta2,theta)
    ELR<-c(ELR2,ELR)
  }
  else if (Lower!=theta[1] && Upper==theta[length(theta)]) {
    u<-u+step
    theta2<-seq(u-step+by,u-by,by)
    ELR2<-rep(NA,length(theta2))
    for (i in 1:length(theta2)) {
      ELR2[i]<-el.cen.EM.G(x,d,fun,theta2[i],err)$'-2LLR'
    }
    theta<-c(theta,theta2)
    ELR<-c(ELR,ELR2)
  }
  else if (Lower==theta[1] && Upper==theta[length(theta)]) {
    l<-l-step
```

```
      theta2<-seq(l+by,l+step-by,by)
      ELR2<-rep(NA,length(theta2))
      for (i in 1:length(theta2)) {
        ELR2[i]<-el.cen.EM.G(x,d,fun,theta2[i],err)$'-2LLR'
      }
      theta<-c(theta2,theta)
      ELR<-c(ELR2,ELR)
      u<-u+step
      theta2<-seq(u-step+by,u-by,by)
      ELR2<-rep(NA,length(theta2))
      for (i in 1:length(theta2)) {
        ELR2[i]<-el.cen.EM.G(x,d,fun,theta2[i],err)$'-2LLR'
      }
      theta<-c(theta,theta2)
      ELR<-c(ELR,ELR2)
    }
    Lower<-min(theta[ELR<=quan])
    Upper<-max(theta[ELR<=quan])
  }
  plot(theta,ELR,cex=0.5,xlab="Theta",ylab="-2logELR",
       main="Empirical Likelihood Ratio Confidence Interval")
  abline(h=3.84)
  abline(v=Lower)
  abline(v=Upper)
  return (c(Lower,Upper))
}
```

```
### The function to generate EM confidence interval using ###
### Qin et al.'s method (2011) ###
findELci2<-function(x,fun1,fun2,alpha,lower,upper,by,step) {
  n<-length(x)
  d<-rep(1,n)
  theta<-seq(lower,upper,by)
  ELR<-rep(NA,length(theta))
  for (i in 1:length(theta)) {
    ELR[i]<-el.cen.EM.G2(x,d,fun1,fun2,theta[i])$adjllr
  }
  Lower<-min(theta[ELR<=qchisq(1-alpha,1)])
  Upper<-max(theta[ELR<=qchisq(1-alpha,1)])
  l<-lower
  u<-upper
  while (Lower==theta[1] || Upper==theta[length(theta)]) {
    if (Lower==theta[1] && Upper!=theta[length(theta)]) {
      l<-l-step
      theta2<-seq(l+by,l+step-by,by)
      ELR2<-rep(NA,length(theta2))
      for (i in 1:length(theta2)) {
        ELR2[i]<-el.cen.EM.G2(x,d,fun1,fun2,theta2[i])$adjllr
      }
      theta<-c(theta2,theta)
      ELR<-c(ELR2,ELR)
    }
    else if (Lower!=theta[1] && Upper==theta[length(theta)]) {
```

```
  u<-u+step
  theta2<-seq(u-step+by,u-by,by)
  ELR2<-rep(NA,length(theta2))
  for (i in 1:length(theta2)) {
    ELR2[i]<-el.cen.EM.G2(x,d,fun1,fun2,theta2[i])$adjllr
  }
  theta<-c(theta,theta2)
  ELR<-c(ELR,ELR2)
}
else if (Lower==theta[1] && Upper==theta[length(theta)]) {
  l<-l-step
  theta2<-seq(l+by,l+step-by,by)
  ELR2<-rep(NA,length(theta2))
  for (i in 1:length(theta2)) {
    ELR2[i]<-el.cen.EM.G2(x,d,fun1,fun2,theta2[i])$adjllr
  }
  theta<-c(theta2,theta)
  ELR<-c(ELR2,ELR)
  u<-u+step
  theta2<-seq(u-step+by,u-by,by)
  ELR2<-rep(NA,length(theta2))
  for (i in 1:length(theta2)) {
    ELR2[i]<-el.cen.EM.G2(x,d,fun1,fun2,theta2[i])$adjllr
  }
  theta<-c(theta,theta2)
  ELR<-c(ELR,ELR2)
}
```

```
    Lower<-min(theta[ELR<=qchisq(1-alpha,1)])
    Upper<-max(theta[ELR<=qchisq(1-alpha,1)])
  }
  plot(theta,ELR,cex=0.5,xlab="Theta",ylab="-2logELR",
       main="Empirical Likelihood Ratio Confidence Interval")
  abline(h=3.84)
  abline(v=Lower)
  abline(v=Upper)
  return (c(Lower,Upper))
}

### The function to generate confidence
### interval using Peng's method (2011) ###
findci<-function(x,sp,alpha,lower,upper,by,step) {
  n<-length(x)
  mid<-n/2
  y<-(x[1:mid]+x[(mid+1):n])/2
  z<-rep(NA,mid)
  for (i in 1:mid) {
    z[i]<-min(x[i],x[mid+i])
  }
  theta<-seq(lower,upper,by)
  ELR<-rep(NA,length(theta))
  for (i in 1:length(theta)) {
    ELR[i]<-llr(theta[i],y,z,sp)
  }
  Lower<-min(theta[ELR<=qchisq(1-alpha,1)])
```

```
Upper<-max(theta[ELR<=qchisq(1-alpha,1)])
l<-lower
u<-upper
while (Lower==theta[1] || Upper==theta[length(theta)]) {
  if (Lower==theta[1] && Upper!=theta[length(theta)]) {
    l<-l-step
    theta2<-seq(l+by,l+step-by,by)
    ELR2<-rep(NA,length(theta2))
    for (i in 1:length(theta2)) {
      ELR2[i]<-llr(theta2[i],y,z,sp)
    }
    theta<-c(theta2,theta)
    ELR<-c(ELR2,ELR)
  }
  else if (Lower!=theta[1] && Upper==theta[length(theta)]) {
    u<-u+step
    theta2<-seq(u-step+by,u-by,by)
    ELR2<-rep(NA,length(theta2))
    for (i in 1:length(theta2)) {
      ELR2[i]<-llr(theta2[i],y,z,sp)
    }
    theta<-c(theta,theta2)
    ELR<-c(ELR,ELR2)
  }
  else if (Lower==theta[1] && Upper==theta[length(theta)]) {
    l<-l-step
    theta2<-seq(l+by,l+step-by,by)
```

```
      ELR2<-rep(NA,length(theta2))
      for (i in 1:length(theta2)) {
        ELR2[i]<-llr(theta2[i],y,z,sp)
      }
      theta<-c(theta2,theta)
      ELR<-c(ELR2,ELR)
      u<-u+step
      theta2<-seq(u-step+by,u-by,by)
      ELR2<-rep(NA,length(theta2))
      for (i in 1:length(theta2)) {
        ELR2[i]<-llr(theta2[i],y,z,sp)
      }
      theta<-c(theta,theta2)
      ELR<-c(ELR,ELR2)
    }
    Lower<-min(theta[ELR<=qchisq(1-alpha,1)])
    Upper<-max(theta[ELR<=qchisq(1-alpha,1)])
  }
  plot(theta,ELR,cex=0.5,xlab="Theta",ylab="-2logELR",
       main="Empirical Likelihood Ratio Confidence Interval")
  abline(h=3.84)
  abline(v=Lower)
  abline(v=Upper)
  return (c(Lower,Upper))
}
```

```
### The function to compare the average
### length of the confidence intervals
### using three methods ###
AVL_Comp<-function(n,m,fun1,fun2,mu,err,alpha,
lower,upper,by,step,sp,dist,cali=FALSE) {
  d<-rep(1,n)
  mid<-n/2
  avl1<-rep(NA,m)
  avl2<-rep(NA,m)
  avl3<-rep(NA,m)
  cp1<-0
  cp2<-0
  cp3<-0
  for (j in 1:m) {
    if (dist=="exp")
       x<-rexp(n,rate=1)
    if (dist=="chisq")
       x<-rchisq(n,df=1)
    if (dist=="lnorm")
       x<-rlnorm(n)

    y<-(x[1:mid]+x[(mid+1):n])/2
    z<-rep(NA,mid)
    for (i in 1:mid) {
      z[i]<-min(x[i],x[mid+i])
    }
```

```
    if (cali)
       quan<-qf(1-alpha,1,n-2)
    else
       quan<-qchisq(1-alpha,1)

    ci1<-findELci(x,fun1,err,alpha,lower,upper,by,step,cali)
    ci2<-findELci2(x,fun1,fun2,alpha,lower,upper,by,step)
    ci3<-findci(x,sp,alpha,lower,upper,by,step)
    if (el.cen.EM.G(x,d,fun1,mu,err)$'-2LLR'<=quan)
      cp1<-cp1+1
    if (el.cen.EM.G2(x,d,fun1,fun2,mu)$adjllr<=qchisq(1-alpha,1))
      cp2<-cp2+1
    if (llr(mu,y,z,sp)<=qchisq(1-alpha,1))
      cp3<-cp3+1
    avl1[j]<-ci1[2]-ci1[1]
    avl2[j]<-ci2[2]-ci2[1]
    avl3[j]<-ci3[2]-ci3[1]
  }
  list(CP1=cp1/m,CP2=cp2/m,CP3=cp3/m,AVL1=mean(avl1),AVL2=mean(avl2)
  ,AVL3=mean(avl3))
}
```

Bibliography

[1] Aalen, O.O. (1976). Nonparametric Inference in Connection with Multiple Decrement Models. *Scand. J. Statist. 3, 15-27.*

[2] Akritas, M.G. (2000). The Central Limit Theorem under Censoring. *Bernoulli, 6, pp. 1109-1120.*

[3] Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes. Springer Verlag, New York.*

[4] Barton, W. (2010). *Comparison of Two Samples by A Nonparametric Likelihood-Ratio Test. Ph.D Dissertation, Department of Statistics, University of Kentucky.*

[5] Breslow, N.E. and Crowley, J.J. (1974). A Large Sample Study of the Life Table and Product Limit Estimates under Random Censorship. *Ann. Statist. 2, 437-453.*

[6] Casella, G. and Berger, R. L. (2002). *Statistical Inference. Duxbury Press, Pacific Grove, California.*

[7] Ceriani, L. and Verme, P. (2011). The Origins of the Gini Index: Extracts from Variabilita e Mutabilita (1912) by Corrado Gini. *J. Econ. Ineq.*

[8] Chen, K and Zhou, M (2007). Computation of the Empirical Likelihood Ratio from Censored Data. *Journal of Statistical Computation and Simulation 77(12):1033-1042.*

[9] Chow, Y.S. and Teicher, H. (1980). *Probability Theory. Springer-Verlag, New York, 1980.*

[10] David, H.A. (1968). Gini's Mean Difference Rediscovered. *Biometrika, 55,573-575.*

[11] Efron, B. (1967). The Two Sample Problem with Censored Data. *In Preceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 4, 831-853.*

[12] Fernholz, L.T. (1983) *von Mises Calculus for Statistical Functionals. Springer-Verlag, New York, Berlin, Heidelberg Tokyo.*

[13] Filippova, A.A. (1962). Mises Theorem on the Asymptotic Behavior of Functionals of Empirical Distribution Function and Its Statistical Applications. *Theory Prob. Appl. 7, 24-57.*

[14] Gentleman, R. and Ihaka, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics, 5, 299-314.*

[15] Gini, C. (1912). *Variabilita e Mutabilita, Contributo Allo Studio Delle Distribuzioni e Relazioni Statistiche. Studi Economico-Giuridici della R. Universita di Cagliari.*

[16] Gill, R.D. (1983). Large Sample Behavior of the Product-Limit Estimator on the Whole Line. *Ann. Statist.,11, 49-58.*

[17] Gill, R.D. (1989). Non- and Semi-Parametric Maximum Likelihood Estimators and the Von Mises Method, Part I. *Scand. J. Statist. 16 (1989), 97-128.*

[18] Glasser, G.J. (1962). Variance Formulas for the Mean Difference and Coefficient of Concentration. *Journal of the American Statistical Association, 57, 648-654.*

[19] Hampel, F.R. (1974). The Influence Curve and Its Role in Robust Estimation. *Journal of American Statistical Assoc. 69, 383-393.*

[20] Hoeffding, W. (1948). A Class of Statistics with Asymptotically Normal Distribution. *Ann. Math. Statist., 19,293-325.*

[21] Hu, Y. (2011). Some Contributions to the Censored Empirical Likelihood With Hazard-Type Constraints. *Ph.D Dissertation, Department of Statistics, University of Kentucky.*

[22] Huber, P.J. (1964). Robust Estimation of A Location Parameter. *Ann. Math. Statist. 35,73-101.*

[23] Kalbfleisch, J.D. and Prentiss, R.L. (2002). *The Statistical Analysis of Failure Time Data. Wiley-Interscience, Hoboken, New Jersey.*

[24] Kaplan, E.L. and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association 53(282):457-481.*

[25] Karagiannis, E. and Kovacevic, M. (2000). A Method to Calculate the Jackknife Variance Estimator for the Gini Coefficient. *Oxford Bulletin of Economics and Statistics, 62, 119-122.*

[26] Langel, M and Tille, Y (2013). Variance Estimation of the Gini index, Revisiting A Result Several Times Published. *J.R.Statist. Soc. A 176,Part1,pp.*

[27] Li, G. (1995). On Nonparametric Likelihood Ratio Estimation Of Survival Probabilities for Censored Data. *Statist. and Prob. Letters, 25, pp. 95-104.*

[28] Lomnicki, Z.A. (1952). The Standard Error of Gini's Mean Difference. *Ann. Math. Statist., 23,635-637.*

[29] Murphy, S.A. (1995). Likelihood Ratio-Based Confidence Intervals in Survival Analysis. *J. Amer. Stat. Assoc., 90, pp. 1399-1405.*

[30] Murphy, S. and van der Vaart, A. (1997). Semiparametric Likelihood Ratio Inference. *Ann. Statist. 25. 1471-1509.*

[31] Nair, U.S. (1936). The Standard Error of Gini's Mean Difference. *Biometrika, 28, 428-436.*

[32] Nelson, W. (1969). Hazard Plotting for Incomplete Failure Data. *J. Qual Technol. 1, 27-52.*

[33] Nelson, W. (1972). Theory and Applications of Hazard Plotting for Censored Failure Data. *Technometrics 14, 945-965.*

[34] Nocedal, J. and Wright, S. (1999). *Numerical Optimization. Springer, New York.*

[35] Owen, A.B. (1988). Empirical Likelihood Ratio Confidence Intervals for A Single Functional. *Biometrika, 75, pp. 237-249.*

[36] Owen, A.B. (1990). Empirical Likelihood Ratio Confidence Regions. *Ann. Stat., 18, pp. 90-120.*

[37] Owen, A.B. (1991). Empirical Likelihood for Linear Models. *Ann. Stat., 19, pp. 1725-1747.*

[38] Owen, A.B. (2001). Empirical Likelihood. *Biometrika, 62(2):269276.*

[39] Pan, X. R. (1997). Empirical Likelihood Ratio Method for Censored Data. *Ph.D Dissertation, Department of Statistics, University of Kentucky.*

[40] Pan, X. R. and Zhou, M. (1999). Using One-parameter Sub-family Of Distributions In Empirical Likelihood With Censored Data. *J. Statist. Planning and Infer., 75, pp. 379-392.*

[41] Pan, X. R. and Zhou, M. (2002). Empirical Likelihood Ratio In Terms Of Cumulative Hazard Function For Censored Data. *. Multivar. Anal., 80, pp. 166-188.*

[42] Peng, L. (2011). Empirical Likelihood Methods for the Gini Index. *Aust. New Zeal. J. Statist.,53,131-139.*

[43] Qin, Y.,Rao, J. and Wu, C. (2010). Empirical Likelihood Confidence Intervals for the Gini Measure of Income Inequality. *Econ. Modeling, 27, 1429-1435.*

[44] Reeds, J.A. (1976). On the Definition of von Mises Functionals. *Ph.D. Dissertation, Harvard University, Cambridge, MA.*

[45] Sandstrom, A., Wretman, J.H. and Walden, B. (1985). Variance Estimators of the Gini coefficient: Simple Random Sampling. *Metron,43,41-70.*

[46] Sandstrom, A., Wretman, J.H. and Walden, B. (1988). Variance Estimators of the Gini coefficient: Probability Sampling. *J. Bus. Econ. Statist., 6,113-120.*

[47] Serfling, R.J. (1980). *Approximation Theorem of Mathematical Statistics. John Wiley & Sons, New York.*

[48] Summers, R. and Heston, A. (1995) The Penn World Tables, Version 5.6. *Cambridge, MA: NBER.*

[49] Thomas, D.R. and Grunkemeier, G.L. (1975). Confidence Interval Estimation Of Survival Probabilities For Censored Data. *J. Amer. Statist. Assoc., 70, pp. 865-871.*

[50] Turnbull, B. (1976). The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data. *Journal of the Royal Statistical Society, Ser. B, 290-295.*

[51] van der Vaart (1997). *Asymptotic Statistics. Cambridge University Press.*

[52] von Mises, R. (1947). On the Asymptotic Distribution of Differentiable Statistical Functions. *Ann. Math. Statist. 18, 309-348.*

[53] Wasserman, L. (2006). *All of Nonparametric Statistics. Springer.*

[54] Xu, K. (2004). How Has the Literature on Gini's Index Evolved in the Past 80 Years?. *Working Paper, Department of Economics, Dalhousie University, Halifax.*

[55] Yitzhaki, S. (1991). Calculating Jackknife Variance Estimators for Parameter of the Gini Method. *Journal of Business* & *Economics Statistics, 9, 235-239.*

[56] Zhou, M. (1991). Some Properties of the Kaplan-Meier Estimator for Independent Nonidentically Distributed Random Variables. *Ann. Stat., 19, pp. 2266-2274.*

[57] Zhou, M. (2005) . Empirical Likelihood Ratio with Arbitrarily Censored/Trucated Data by EM Algorithm. *Journal of Computational and Graphical Statistics 14(3):643-656.*

[58] Zhou, M. and Yang, Y. (2015). A Recursive Formula for the Kaplan-Meier Estimator with Mean Constraints and Its Application to Empirical Likelihood. *Computational Statistics.*

[59] Zhou, M. (2015). *Empirical Likelihood Method in Survival Analysis. CRC Press.*

Vita

ZHIYUAN SHEN

EDUCATION

Ph.D., Statistics, University of Kentucky, Lexington, KY **2010-present**
M.S., Statistics, University of Kentucky, Lexington, KY **2010-2012**
B.S., Math. and Applied Math., Peking University, Beijing, China **2006-2010**
B.S., Economics, Peking University, Beijing, China **2006-2010**

EXPERIENCE

Research Assistant,University of Kentucky, Lexington, KY, **2012-present**
Teaching Assistant, University of Kentucky, Lexington, KY, **2010-2012**

CERTIFICATES

Passed Charted Financial Analyst(CFA) Level 2 exam **2013**
Passed Charted Financial Analyst(CFA) Level 1 exam **2012**
SAS Certified Advanced Programmer for SAS 9 Credential **2013**
SAS Certified Base Programmer for SAS 9 Credential **2011**

ACTIVITIES & HONORS

Website Manager, Statistics Students Association, University of Kentucky, **2012-2013**
Second Prize, Competition of Simulating Stock Market attended by hundreds of students from universities in Beijing. **2007**