



University of Kentucky
UKnowledge

University of Kentucky Doctoral Dissertations

Graduate School

2011

SOME CONTRIBUTIONS TO THE CENSORED EMPIRICAL LIKELIHOOD WITH HAZARD-TYPE CONSTRAINTS

Yanling Hu

University of Kentucky, vivianhyl@hotmail.com

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

Recommended Citation

Hu, Yanling, "SOME CONTRIBUTIONS TO THE CENSORED EMPIRICAL LIKELIHOOD WITH HAZARD-TYPE CONSTRAINTS" (2011). *University of Kentucky Doctoral Dissertations*. 150.
https://uknowledge.uky.edu/gradschool_diss/150

This Dissertation is brought to you for free and open access by the Graduate School at UKnowledge. It has been accepted for inclusion in University of Kentucky Doctoral Dissertations by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

ABSTRACT OF DISSERTATION

Yanling Hu

The Graduate School
University of Kentucky
2011

SOME CONTRIBUTIONS TO THE CENSORED EMPIRICAL LIKELIHOOD
WITH HAZARD-TYPE CONSTRAINTS

ABSTRACT OF 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
Yanling Hu
Lexington, Kentucky

Director: Dr. Mai Zhou, Professor of Statistics
Lexington, Kentucky 2011

Copyright© Yanling Hu 2011

ABSTRACT OF DISSERTATION

SOME CONTRIBUTIONS TO THE CENSORED EMPIRICAL LIKELIHOOD WITH HAZARD-TYPE CONSTRAINTS

Empirical likelihood (EL) is a recently developed nonparametric method of statistical inference. Owen's 2001 book contains many important results for EL with uncensored data. However, fewer results are available for EL with right-censored data. In this dissertation, we first investigate a right-censored-data extension of Qin and Lawless (1994). They studied EL with uncensored data when the number of estimating equations is larger than the number of parameters (over-determined case). We obtain results similar to theirs for the maximum EL estimator and the EL ratio test, for the over-determined case, with right-censored data. We employ hazard-type constraints which are better able to handle right-censored data. Then we investigate EL with right-censored data and a k -sample mixed hazard-type constraint. We show that the EL ratio test statistic has a limiting chi-square distribution when $k = 2$. We also study the relationship between the constrained Kaplan-Meier estimator and the corresponding Nelson-Aalen estimator. We try to prove that they are asymptotically equivalent under certain conditions. Finally we present simulation studies and examples showing how to apply our theory and methodology with real data.

KEYWORDS: Asymptotic Chi-square Distribution, Censored Data, Empirical Likelihood, Martingale Theorem, Over-determined Constraints

Author's signature: Yanling Hu

Date: April 11, 2011

RULES FOR THE USE OF DISSERTATIONS

Unpublished dissertations submitted for the Doctor's degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgments.

Extensive copying or publication of the dissertation in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

A library that borrows this dissertation for use by its patrons is expected to secure the signature of each user.

Name

Date

DISSERTATION

Yanling Hu

The Graduate School
University of Kentucky
2011

SOME CONTRIBUTIONS TO THE CENSORED EMPIRICAL LIKELIHOOD
WITH HAZARD-TYPE CONSTRAINTS

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
Yanling Hu
Lexington, Kentucky

Director: Dr. Mai Zhou, Professor of Statistics
Lexington, Kentucky 2011

Copyright© Yanling Hu 2011

To my dear husband Piao, and my parents.

ACKNOWLEDGMENTS

First of all, I would like to express my gratitude to my advisor, Dr. Mai Zhou, whose encouragement, guidance, and patience have made this dissertation possible.

I would also like to thank my dissertation committee, Drs. Arne Bathke, Richard Charnigo, Yanbing Zheng and Robert Molzon, for their advice and helpful comments.

Finally I dedicate this dissertation to my parents, Mr. Anguo Hu and Ms. Suxin Ji, and to my husband Piao Liu, for their encouragement and support of my study.

TABLE OF CONTENTS

Acknowledgments	ix
Table of Contents	x
List of Figures	xii
Chapter 1 Outline of the Dissertation	1
Chapter 2 Introduction	3
2.1 Empirical Likelihood Ratio Test for Uncensored Data	3
2.2 Empirical Likelihood Ratio Test for Censored Data	7
Chapter 3 Censored Empirical Likelihood with Over-determined Hazard-type Constraints	12
3.1 Background	12
3.2 Empirical Likelihood, Over-determined Constraints in Terms of Hazard	14
3.3 Asymptotic Results for Empirical-Likelihood-Ratio Statistic and Maximum- Empirical-Likelihood Estimator	18
3.4 Lemmas and Proofs	25
3.5 Computational Considerations	41
Chapter 4 Hazard-Type Hypothesis for K-sample U-statistic	44
4.1 Introduction	44
4.2 MLE's for the Hazard Jumps	48
4.3 ELLR for the Two Samples	59
4.4 Distribution of -2ELLR	63
4.5 Validation of Taylor Expansions	69
4.6 Lemmas	72
Chapter 5 Relationship between Constrained Kaplan-Meier Estimators and the Corresponding Nelson-Aalen Estimators	77
5.1 Background	77
5.2 Relationship Between the Constrained Kaplan-Meier Estimator and the Corresponding Nelson-Aalen Estimator	80
Chapter 6 Simulations and Examples	87
6.1 Simulations	87
6.2 Examples	98
Chapter 7 Discussions and Future Research Questions	102
Appendix: Program Code	103

Bibliography	110
Vita	114

LIST OF FIGURES

6.1	Simulation-sample size 50 with true parameter	88
6.2	Simulation-sample size 100 with true parameter	88
6.3	Simulation-with censoring distribution: 20+Exponential ($\lambda_1 = 0.02$) . . .	89
6.4	Simulation-sample size 50 with false parameter	90
6.5	Empirical likelihood ratio statistic with different constraints, n=100 . . .	92
6.6	Empirical likelihood ratio statistic with different constraints, n=100 . . .	92
6.7	Empirical likelihood ratio statistic with different constraints, n=50	93
6.8	Empirical likelihood ratio statistic with different constraints, n=50	93
6.9	Empirical likelihood ratio statistic with different constraints, n=20	94
6.10	Empirical likelihood ratio statistic with different constraints, n=20	94
6.11	Empirical likelihood ratio statistic with different constraints and censoring distribution: 20+Exponential ($\lambda_1 = 0.02$), n=100	95
6.12	Empirical likelihood ratio statistic with different constraints and censoring distribution: 20+Exponential ($\lambda_1 = 0.02$), n=100	95
6.13	Empirical likelihood ratio statistic with different constraints and censoring distribution: 20+Exponential ($\lambda_1 = 0.02$), n=50	96
6.14	Empirical likelihood ratio statistic with different constraints and censoring distribution: 20+Exponential ($\lambda_1 = 0.02$), n=50	96
6.15	Empirical likelihood ratio statistic with different constraints and censoring distribution: 20+Exponential ($\lambda_1 = 0.02$), n=20	97
6.16	Empirical likelihood ratio statistic with different constraints and censoring distribution: 20+Exponential ($\lambda_1 = 0.02$), n=20	97

Chapter 1 Outline of the Dissertation

This dissertation is organized as follows:

In Chapter 2, we review the empirical-likelihood-ratio test for both uncensored and censored data.

In Chapter 3, we discuss some results for empirical likelihood and general estimating equations developed by Qin and Lawless (1994) [30]. Qin and Lawless worked with uncensored data and used distribution-type empirical likelihood and constraint/estimating equations. We present a parallel construct to that of Qin and Lawless, using a hazard-type empirical likelihood with over-determined hazard-type estimating equations/constraints with right-censored data. This approach naturally incorporates the censoring, and the empirical-likelihood estimator and test statistic also have nice asymptotic properties, similar to what Qin and Lawless obtained.

In Chapter 4, we present a mathematical derivation for the case where a two-sample mixed hazard-type hypothesis is used. This chapter is a collaboration with Dr. Bill Barton. We believe that the analysis of this hazard-type hypothesis is a valuable theoretical contribution in its own right. In addition we believe it represents another step after Chapter 3 toward establishing a relationship between distribution-type hypothesis and hazard-type hypothesis for right-censored data. The martingale central limit theorem is used. We show that the empirical-likelihood-ratio statistic has an asymptotic $\chi^2_{(1)}$ distribution.

In Chapter 5, we study the relationship between the constrained Kaplan-Meier

estimator and the corresponding constrained Nelson-Aalen estimator under certain conditions. Akritas (2000) [1] established a relationship between the Kaplan-Meier estimator and the corresponding Nelson-Aalen estimator. Such a relationship is valuable since hazard-type hypotheses are typically more mathematically tractable than distribution-type hypotheses, for censored data. So we hope to get a further relationship between a general constrained Kaplan-Meier estimator and the corresponding constrained Nelson-Aalen estimator and we argue that once such relationship is constructed in a certain format, many existing distribution-type empirical-likelihood results can be converted to hazard-type empirical-likelihood results for handling censored data.

Simulation studies and examples showing how to implement the theory in Chapter 3 are presented and discussed in Chapter 6.

Chapter 7 summarizes our findings and discusses possible future investigations.

Finally, the R code for the simulations is listed in the Appendix.

Chapter 2 Introduction

2.1 Empirical Likelihood Ratio Test for Uncensored Data

The method of maximum likelihood is one of the most popular techniques for deriving estimators. The likelihood-ratio tests are widely used for the maximum likelihood estimation. Let us consider the parametric likelihood-ratio test (LRT) first.

Suppose that X_1, X_2, \dots, X_n are i.i.d. random variables from a population with pdf or pmf $f(x|\theta_1, \dots, \theta_k)$, and x_1, x_2, \dots, x_n are the corresponding observations. The likelihood function is defined by

$$L(\boldsymbol{\theta}|\mathbf{x}) = L(\theta_1, \dots, \theta_k|x_1, \dots, x_n) = \prod_{i=1}^n f(x_i|\theta_1, \dots, \theta_k). \quad (2.1)$$

The likelihood-ratio test statistic for testing $H_0 : \boldsymbol{\theta} \in \Theta_0$ versus $H_1 : \boldsymbol{\theta} \in \Theta_0^c$ is defined as

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

where Θ denotes the full parameter space, Θ_0 is some subset of Θ , and Θ_0^c is its complement. We reject H_0 when $\lambda(\mathbf{x})$ is less than some threshold value c , where $0 < c < 1$. Wilks (1938) [32] showed that when the null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ is true, then the test statistic $-2 \log \lambda(\mathbf{x})$ has an asymptotic $\chi_{(p)}^2$ distribution under certain regularity conditions, where p is the number of restrictions imposed on the parameters by H_0 .

The parametric LRT is useful for finding efficient estimators, constructing tests, and finding confidence intervals. However, it is only applicable when we know what the parametric family is. If we do not know what the parametric family is then we may use an empirical likelihood-ratio test (ELRT), which is a LRT without distribu-

tional assumptions.

The ELRT is one of the most useful nonparametric methods for statistical inference. It can be used to conduct hypothesis tests and to find confidence intervals in ways that are analogous to those of the parametric LRT, but which make no strong distributional assumptions. Let us now consider the ELRT.

Let X_1, X_2, \dots, X_n be i.i.d. random variables with an unknown distribution F_0 , and let x_1, x_2, \dots, x_n be the corresponding observations. The empirical likelihood function based on the observations is defined as

$$\begin{aligned} L(F) &= \prod_{i=1}^n [F(x_i) - F(x_{i-})] \\ &= \prod_{i=1}^n dF(x_i) \\ &= \prod_{i=1}^n p_i, \end{aligned} \tag{2.3}$$

where

$$p_i = dF(x_i) = F(x_i) - F(x_{i-}), \tag{2.4}$$

and $L(F)$ is the probability of getting exactly the observed sample values x_1, \dots, x_n from the cumulative distribution function F .

It can be shown that the empirical cumulative distribution function (ECDF), or empirical distribution, maximizes $L(F)$ over all possible distribution functions (Owen (2001) [26]). The ECDF for x_1, x_2, \dots, x_n is defined as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(x_i \leq x), \quad \text{for } -\infty < x < \infty. \tag{2.5}$$

Thomas and Grunkemeier (1975) [31] first investigated the ELRT for randomly-censored data. They proposed empirical-likelihood-ratio test methods for confidence-interval estimation of survival or life-time probabilities for randomly censored data.

They showed heuristically that the empirical likelihood ratio for a survival probability has a limiting $\chi^2_{(1)}$ distribution under the constraint $P(X > a) = p$, where a is a real number and p is a hypothesized probability.

Building on the preceding proposition of Thomas and Grunkemeier (1975), Owen (1988, 1990) [22] [23] developed the ELRT for constructing confidence intervals and tests for uncensored data.

The empirical-likelihood-ratio function for i.i.d. uncensored data is defined as

$$\begin{aligned} R(F) &= \frac{L(F)}{L(F_n)} \\ &= \prod_{i=1}^n np_i. \end{aligned} \tag{2.6}$$

Note that this formula does not require that X_i 's be distinct. We reject H_0 when $R(F)$ is less than some threshold value c , where $0 < c < 1$.

Suppose that we are interested in a parameter $\theta = T(F)$, where T is a given real-valued function. For simplicity, we consider the mean of F , so $\theta = \int x dF(x)$. To make an inference about θ , we follow Owen and define the profile empirical likelihood ratio function

$$R(\theta) = \sup_F \{R(F) \mid \sum_{i=1}^n p_i x_i = \theta, p_i \geq 0, \sum_{i=1}^n p_i = 1\}. \tag{2.7}$$

Owen has shown that under the null hypothesis: $T(F_0) = \theta_0$, the empirical-likelihood-ratio statistic $-2 \log R(\theta_0)$ has an asymptotic $\chi^2_{(1)}$ distribution. Thus the empirical-likelihood-ratio method developed by Owen can be used analogously to the likelihood-ratio method in parametric settings, but without strong distribution assumptions. Hypothesis tests and confidence intervals can be similarly obtained. Later, Owen (1991) [24] made extensions to regression problems, and Kolaczyk (1992) [16] and Owen (1992) [25] made further extensions to generalized linear models and projec-

tion pursuit regression.

Other asymptotic properties of empirical-likelihood-ratio statistics have been studied by DiCiccio, Romano, Hall, and many others. Parametric and empirical likelihood functions or surfaces were compared by DiCiccio, Hall and Romano (1989) [5]. DiCiccio and Romano (1989) [7] considered the standard multivariate-normal approximation to the distribution of the signed root of the empirical-likelihood-ratio statistic in cases where inference is required for a smooth function of the mean of the distribution from which the sample is drawn. Hall (1990) [10] proved that, except for a location term, empirical likelihood draws contours which are second-order correct for those of a pseudo-likelihood. DiCiccio, Hall, and Romano (1991) [6] showed that in a very general setting, the empirical likelihood method for constructing confidence intervals is Bartlett-correctable. That result makes empirical likelihood competitive with methods such as the bootstrap which are not Bartlett-correctable, and most importantly, demonstrates a strong link between empirical likelihood and parametric likelihood, since the Bartlett correction had previously only been available for parametric likelihood.

Qin and Lawless (1994) [30] studied empirical likelihood and general estimating equations for uncensored data. They showed that the empirical-likelihood method could be naturally brought to bear on problems with over-determined estimating equations, where the number of estimating equations r is greater than the number of parameters p . They demonstrated how the maximum-empirical-likelihood estimators of parameters $\boldsymbol{\theta} \in \mathfrak{R}^p$ may be obtained and they determined the asymptotic multivariate-normal distribution for the estimators. They also proved that the empirical-likelihood-ratio test statistic for the parameters has an asymptotic $\chi_{(p)}^2$ distribution, hence confidence regions and hypothesis tests can be constructed. They

also showed that the maximum-empirical-likelihood estimator is “efficient”.

2.2 Empirical Likelihood Ratio Test for Censored Data

In this section, we consider the application of the ELRT to censored data. We begin by describing the three types of censored data, which are right-censored data, left-censored data, and interval-censored data. To this end we cite the description in Chen (2005) [4]:

Among these censoring mechanisms, right-censoring is the most common. An observation on a variable T is right censored if all you know about T is that it is greater than some value c . In survival analysis, T is typically the time of occurrence for some event, and cases are right censored because observation of events is terminated at c before the event occurs. For example, in a five-year study of mortality from lung cancer, survival time will be right censored for patients who are still alive at the end of the five year period.

Left censoring occurs when the only information you know about an observation on a variable T is that it is less than some value. In the context of survival data, left censoring is the most likely to occur when you begin observing a sample at a time when some of the individuals may have already experienced the event. For example, in a study of times of occurrence of developmental milestones in children, some children already have achieved the milestones prior to enter into the study.

Interval censoring combines both right and left censoring. An observa-

tion on a variable T is interval censored if all you know about T is that $a < T < b$, for some values of a and b . Interval censoring occurs when the patient in a clinical trial or longitudinal study has periodic follow-up and the patient's event time is only known to fall in an interval. For instance, a sample of people is tested annually for HIV infection. If a person who was not infected at the end of year two is then found to be infected at the end of year three, the time of infection is interval censored between two and three. This type of censoring may also occur in industrial experiment where there is periodic inspection for proper functioning of equipment items.

For censored data it is mathematically convenient to write the likelihood in terms of hazard functions. Murphy (1995) [20] discussed two extensions to the empirical likelihood, the Poisson extension and the Binomial extension, for estimation of survival probability. She showed that the likelihood-ratio statistics of both extensions have limiting $\chi^2_{(1)}$ distributions, for the constraint $\Lambda(t_0) = \theta_0$, where $\Lambda(t_0)$ is the cumulative hazard function at t_0 . Murphy and Van der Vaart (1997) [21] also considered empirical-likelihood-ratio tests and related confidence intervals. Their paper includes an empirical likelihood result for $\theta = \int g(t)dF(t)$ with doubly-censored data. But their regularity conditions are too strong.

Pan and Zhou (1999) [28], desiring to enlarge the possible parameters that empirical likelihood can deal with, showed that the empirical likelihood in terms of hazard function for right-censored data has a limiting $\chi^2_{(1)}$ distribution under the constraint

$$\int g(t)d\Lambda(t) = \theta,$$

where $g(t)$ is a given function that satisfies some moment conditions and θ is a given constant. This can be generalized to k constraints as

$$\int g_k(t)d\Lambda_k(t) = \theta_k, \quad k = 1, \dots, p$$

and in this case the empirical-likelihood-ratio statistic has an asymptotic $\chi_{(p)}^2$ distribution.

Building on the preceding proposition of Pan and Zhou, Fang (2000) [8] studied the binomial extension of the empirical likelihood under the constraint

$$\sum_{i=1}^{n-1} g(x_i) \log(1 - \Delta\Lambda(x_i)) = \theta$$

for one-sample and two-sample right-censored data, where $g(\cdot)$ is a given left-continuous weight function and θ is a given constant. Fang demonstrated that the binomial extension of the empirical-likelihood-ratio statistic for censored data has a limiting $\chi_{(1)}^2$ distribution. Similarly, she generalized the result to p constraints

$$\sum_{i=1}^{n-1} g_k(x_i) \log(1 - \Delta\Lambda(x_i)) = \theta_k, \quad k = 1, \dots, p$$

and showed that in this case the empirical-likelihood-ratio statistic has an asymptotic $\chi_{(p)}^2$ distribution.

Zhou (2000) [34] proposed the empirical-envelope-likelihood method for estimation problems. He noted that in some cases, the nonparametric maximum-likelihood estimator does not exist or there is more than one maximizer of the empirical likelihood. One example is the problem of defining the empirical likelihood for the symmetric distribution of location-shift distribution. Zhou proposed to first enlarge the parameter space to make the maximum-empirical-likelihood estimator well-defined and then to gradually shrink the enlarged parameter space by placing more and more restrictions on it. He discussed several problems where this method can be

applied effectively. He also obtained the asymptotic distribution of the empirical-envelope maximum-likelihood-estimator. Later, Kim and Zhou (2004) [15] applied the empirical-envelope-likelihood method to the symmetric-location problem. They showed that the usual asymptotic theory of empirical likelihood still holds and that the asymptotic efficiency of the empirical maximum-likelihood estimator of location is obtained.

As we mentioned above, when there are censored data, it is difficult to get the estimate of the distribution function that maximizes the empirical likelihood under the constraints. Hence, we propose to study censored empirical likelihood with hazard-type constraints, since we can easily get the estimate of the hazard function when we use hazard-type empirical likelihood with hazard-type constraints. We will show that hypothesis tests and confidence intervals can also be developed.

Suppose that x_1, x_2, \dots, x_n are i.i.d. nonnegative observations denoting the lifetimes from a continuous distribution function F_0 . Independent of the lifetimes there are censoring times c_1, c_2, \dots, c_n which are i.i.d. from a distribution G_0 . In practice F_0 and G_0 will be unknown. Only the censored observations, (t_i, δ_i) , are available to us:

$$t_i = \min(x_i, c_i) \quad \text{and} \quad \delta_i = I[x_i \leq c_i] \quad \text{for } i = 1, 2, \dots, n \quad (2.8)$$

The censored empirical likelihood we will study in this dissertation is the one in Pan and Zhou (2002) [29], which was called a Poisson extension of the likelihood by

Murphy (1995) [20]:

$$\begin{aligned} AL(\Lambda) &= \prod_{i=1}^n [\Delta\Lambda(t_i)]^{\delta_i} \exp\{-\Lambda(t_i)\} \\ &= \prod_{i=1}^n [\Delta\Lambda(t_i)]^{\delta_i} \exp\left\{-\sum_{j:t_j \leq t_i} \Delta\Lambda(t_j)\right\} \end{aligned}$$

where $\Lambda(\cdot)$ is the cumulative hazard function and $\Delta\Lambda(t)$ is the jump of Λ at t .

The hazard-type constraint we will investigate is similar to that described in Pan and Zhou (2002):

$$\int g(t) d\Lambda(t) = \theta,$$

where $g(t)$ is a given function that satisfies some moment conditions and θ is a given constant.

Chapter 3 Censored Empirical Likelihood with Over-determined Hazard-type Constraints

3.1 Background

Qin and Lawless (1994) [30] studied empirical likelihood and general estimating equations for uncensored data. They linked estimating functions or equations and empirical likelihood, and also developed methods of combining information about parameters. They did this by assuming that information about distribution function F and parameter vector $\boldsymbol{\theta}$ is available in the form of unbiased estimating/constraint equations.

Let X_1, X_2, \dots, X_n be i.i.d. random variables with an unknown distribution function F , and a p -dimensional parameter $\boldsymbol{\theta}$ associated with F . We assume that information about $\boldsymbol{\theta}$ and F is available in the form of $r \geq p$ functionally independent unbiased estimating functions, that is functions $g_j(X, \boldsymbol{\theta}), j = 1, 2, \dots, r$, such that $E_F g_j(X, \boldsymbol{\theta}) = 0$. This can be represented in vector form as

$$\mathbf{g}(X, \boldsymbol{\theta}) = (g_1(X, \boldsymbol{\theta}), \dots, g_r(X, \boldsymbol{\theta}))^\top,$$

where

$$E_F\{\mathbf{g}(X, \boldsymbol{\theta})\} = \mathbf{0}.$$

Let x_1, x_2, \dots, x_n be i.i.d. observations from F . Then the empirical likelihood function using (2.3) is:

$$L(F) = \prod_{i=1}^n dF(x_i) = \prod_{i=1}^n p_i, \quad (3.1)$$

where

$$p_i = dF(x_i) = Pr(X = x_i).$$

Only distributions with a positive probability on each x_i have non-zero likelihood, and (3.1) is maximized by the empirical distribution function $F_n(x) = \frac{1}{n} \sum_{i=1}^n I(x_i \leq x)$. Then from (2.6) the empirical-likelihood ratio is

$$R(F) = L(F)/L(F_n) = \prod_{i=1}^n np_i.$$

Suppose the parameter $\boldsymbol{\theta}$ is such that $\int \mathbf{g}(x, \boldsymbol{\theta}) dF = \mathbf{0}$. To obtain an inference about $\boldsymbol{\theta}$, we define the profile empirical-likelihood-ratio function

$$R_E(\boldsymbol{\theta}) = \sup \left\{ \prod_{i=1}^n np_i \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \mathbf{g}(x_i, \boldsymbol{\theta}) = \mathbf{0} \right\}. \quad (3.2)$$

As noted by Qin and Lawless (1994), the maximum of $\prod_{i=1}^n np_i$ subject to the constraints

$$p_i \geq 0, \quad \sum_i p_i = 1, \quad \sum_i p_i \mathbf{g}(x_i, \boldsymbol{\theta}) = \mathbf{0}$$

can be found via Lagrange multipliers.

Notice here the number of constraint equations r is greater than or equal to the number of parameters p . When $r = p$ (just-determined case), Qin and Lawless's results are the same as those of Owen (1988, 1990) [22] [23]. However, their main contribution is the case where $r > p$, which we refer to as the case of over-determined constraints.

Qin and Lawless derived the maximum empirical likelihood estimator (MELE) $\tilde{\boldsymbol{\theta}}$, where $\tilde{\boldsymbol{\theta}}$ is the $\boldsymbol{\theta}$ value that achieves the maximum of the empirical likelihood function under those over-determined constraint equations. They also proved that under the null hypothesis: $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, the empirical-likelihood-ratio statistic $W_E(\boldsymbol{\theta}_0) = -2(L_E(\boldsymbol{\theta}_0) - L_E(\tilde{\boldsymbol{\theta}}))$ converges to a $\chi_{(p)}^2$ distribution as $n \rightarrow \infty$, where $L_E(\boldsymbol{\theta}) = \max_{\{p_i: p_i \geq 0, \sum_i p_i = 1, \sum_i p_i \mathbf{g}(x_i, \boldsymbol{\theta}) = \mathbf{0}\}} \sum_{i=1}^n \log p_i(\boldsymbol{\theta})$.

However, Qin and Lawless's results are limited to uncensored data. For right-censored data, no results are available. Hence, we propose a parallel construct to that of Qin and Lawless, which uses a hazard-type empirical likelihood with over-determined hazard-type estimating equations/constraints. This approach naturally incorporates censoring and the empirical likelihood estimator and test statistic also have nice asymptotic properties.

The over-determined constraint problems often appear in econometrics, where people now use the generalized method of moments (GMM) to solve such problems. The GMM is very popular in econometrics when the number of constraint equations is larger than the number of coefficient parameters in a regression model. To use this method, people first need to choose a variance-covariance matrix as the weight matrix to minimize the corresponding combination of the constraints. However, the variance-covariance matrix is usually not efficient and people need to do iterations to update the matrix and make it efficient finally. While using empirical likelihood method to solve these problems, Qin and Lawless proved that their variance-covariance matrix is efficient and no iteration is needed. And the GMM is hardly used at all outside of econometrics.

3.2 Empirical Likelihood, Over-determined Constraints in Terms of Hazard

Suppose that x_1, x_2, \dots, x_n are i.i.d. non-negative observations of lifetimes with a continuous distribution function F_0 . Independent of the lifetimes there are censoring times c_1, c_2, \dots, c_n which are i.i.d. with a distribution G_0 . In practice F_0 and G_0 will be unknown. Only the censored observations, (t_i, δ_i) , are available to us:

$$t_i = \min(x_i, c_i) \quad \text{and} \quad \delta_i = I[x_i \leq c_i] \quad \text{for } i = 1, 2, \dots, n \quad (3.3)$$

The empirical likelihood based on censored observations (t_i, δ_i) pertaining to the distribution function F is defined by (Pan and Zhou (2002) [29]):

$$EL(F) = \prod_{i=1}^n [\Delta F(t_i)]^{\delta_i} [1 - F(t_i)]^{1-\delta_i}.$$

The cumulative hazard function $\Lambda(t)$ related to CDF $F(t)$ is defined by

$$\Lambda(t) = \int_{[0,t)} \frac{dF(s)}{1 - F(s-)}.$$

We will restrict our analysis of the empirical likelihood to the purely discrete functions dominated by their NPMLE's. See Owen(1988) for discussion on this restriction.

In that case the above relation between $\Lambda(\cdot)$ and $F(\cdot)$ gives

$$1 - F(t) = \prod_{s \leq t} (1 - \Delta\Lambda(s)) \quad \text{and} \quad \Delta\Lambda(t) = \frac{\Delta F(t)}{1 - F(t-)}.$$

This in turn allows us to write the empirical likelihood above for discrete $F(\cdot)$ in terms of Λ as:

$$EL(\Lambda) = \prod_{i=1}^n [\Delta\Lambda(t_i)]^{\delta_i} \left[\prod_{j:t_j < t_i} (1 - \Delta\Lambda(t_j)) \right]^{\delta_i} \left[\prod_{j:t_j \leq t_i} (1 - \Delta\Lambda(t_j)) \right]^{1-\delta_i},$$

where $\Delta\Lambda(t) = \Lambda(t+) - \Lambda(t-)$ is the jump of Λ at t . The reason we use $EL(\Lambda)$ instead of $EL(F)$ is that the hazard function naturally incorporates censoring and the estimates are easier to deal with.

The above is called a binomial version of the empirical likelihood. In this dissertation, we will use a simpler version of the $EL(\Lambda)$, which was called a Poisson extension of the likelihood by Murphy (1995) [20] and which was also used by Pan and Zhou (2002):

$$\begin{aligned} AL(\Lambda) &= \prod_{i=1}^n [\Delta\Lambda(t_i)]^{\delta_i} \exp\{-\Lambda(t_i)\} \\ &= \prod_{i=1}^n [\Delta\Lambda(t_i)]^{\delta_i} \exp\left\{-\sum_{j:t_j \leq t_i} \Delta\Lambda(t_j)\right\} \end{aligned}$$

Notice we have assumed a discrete $\Lambda(t)$ in the above. The difference between $AL(\Lambda)$ and $EL(\Lambda)$ is small and negligible for large n . See Pan and Zhou (2002) for the comparison. We also want to mention here that we study the large sample properties of the empirical likelihood statistic and estimators, so $n \gg r$, and r is a fixed positive integer. When $r \rightarrow \infty$, similar results might also hold, but need further investigation.

Let $w_i = \Delta\Lambda(t_i)$ for $i = 1, 2, \dots, n$. Here, without loss of generality, we assume the t_i s are already sorted in an increasing order. The likelihood of this Λ can be written in terms of the jumps

$$AL = \prod_{i=1}^n [w_i]^{\delta_i} \exp\left\{-\sum_{j=1}^n w_j I[t_j \leq t_i]\right\}$$

and the log likelihood is

$$\begin{aligned} \log AL &= \sum_{i=1}^n \left\{ \delta_i \log w_i - \sum_{j=1}^n w_j I[t_j \leq t_i] \right\} \\ &= \sum_{i=1}^n \delta_i \log w_i - \sum_{i=1}^n w_i R_i, \end{aligned} \tag{3.4}$$

where $R_i = \sum_j I[t_j \geq t_i]$.

If we maximize the log AL over all possible hazard functions it is well-known that this yields $\hat{w}_i = \frac{\delta_i}{R_i}$. This is the well-known Nelson-Aalen estimator: $\Delta\hat{\Lambda}_{NA}(t_i) = \frac{\delta_i}{R_i}$.

Next, we want to maximize the log AL with w_i satisfying some estimating equations. To that end let us first discuss the estimating equations in terms of hazard. Denote the parameter vector as

$$\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top.$$

We assume information about $\boldsymbol{\theta}$ and Λ is given either implicitly in equations

$$\left\{ \int g_1(t, \boldsymbol{\theta}) d\Lambda(t), \dots, \int g_r(t, \boldsymbol{\theta}) d\Lambda(t) \right\}^\top = (k_1, \dots, k_r)^\top, \quad (3.5)$$

or, explicitly in equations

$$\left\{ \int g_1(t) d\Lambda(t), \dots, \int g_r(t) d\Lambda(t) \right\}^\top = (f_1(\boldsymbol{\theta}), \dots, f_r(\boldsymbol{\theta}))^\top, \quad (3.6)$$

where $g_i(t, \cdot), i = 1, \dots, r$ are functionally independent estimating functions defined similarly as in Qin and Lawless, and also satisfy some conditions discussed later in the lemmas and theorems; and $f_i(\cdot)$ are some finite functions of $\boldsymbol{\theta}$. Here the number of equations r can be larger than p . In this dissertation, we focus on results for estimators $\hat{\boldsymbol{\theta}}$ of parameters $\boldsymbol{\theta}$ defined as in (3.5) above. The other type of $\boldsymbol{\theta}$ defined by equation (3.6) can be similarly studied and parallel results are seen to also hold. Denote

$$\mathbf{g}(t, \boldsymbol{\theta}) = (g_1(t, \boldsymbol{\theta}), \dots, g_r(t, \boldsymbol{\theta}))^\top,$$

$$\mathbf{k} = (k_1, \dots, k_r)^\top.$$

Example: The parameter θ of median can be defined as $\int I_{[t \leq \theta]} d\Lambda(t) = \log 2$, where $g(t, \theta) = I_{[t \leq \theta]}$, and $k = \log 2$. Other quantiles can be similarly defined.

Notice that $w_n = \delta_n$, because the last jump of a discrete cumulative hazard function must be one. The estimating equations in terms of hazard for the above parameters $\boldsymbol{\theta}$ defined in (3.5) are, by discreteness,

$$\sum_{i=1}^{n-1} \delta_i \mathbf{g}(t_i, \boldsymbol{\theta}) \hat{w}_i + \mathbf{g}(t_n, \boldsymbol{\theta}) \delta_n = \mathbf{k},$$

where $\hat{w}_i = \Delta \hat{\Lambda}(t_i)$.

3.3 Asymptotic Results for Empirical-Likelihood-Ratio Statistic and Maximum-Empirical-Likelihood Estimator

The next step in our empirical likelihood analysis is to find a (discrete) cumulative hazard function w_i that maximizes $\log AL$ under the constraints

$$\sum_{i=1}^{n-1} \delta_i \mathbf{g}(t_i, \boldsymbol{\theta}) w_i + \mathbf{g}(t_n, \boldsymbol{\theta}) \delta_n = \mathbf{k} . \quad (3.7)$$

If $r = p$, the maximum empirical likelihood estimator $\hat{\boldsymbol{\theta}}$ may be obtained as roots of the corresponding hazard-type estimating equations defined above with $\hat{w}_i = \delta_i/R_i$. However, the constrained maximizer w_i in general does not equal the Nelson-Aalen jump when $r > p$.

Theorem 1 *If the constraints (3.7) are feasible (which means there is at least a genuine hazard w_i that solves (3.7)), then the maximum of AL under the constraints is obtained when*

$$\begin{aligned} w_i &= w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) \\ &= \frac{\delta_i}{R_i + n \boldsymbol{\lambda}(\boldsymbol{\theta})^\top \mathbf{g}(t_i, \boldsymbol{\theta}) \delta_i} \\ &= \frac{\delta_i}{R_i} \times \frac{1}{1 + \boldsymbol{\lambda}(\boldsymbol{\theta})^\top (\delta_i \mathbf{g}(t_i, \boldsymbol{\theta}) / (R_i/n))} \\ &= \Delta \hat{\Lambda}_{NA}(t_i) \frac{1}{1 + \boldsymbol{\lambda}(\boldsymbol{\theta})^\top \mathbf{Z}(t_i, \boldsymbol{\theta})} , \end{aligned} \quad (3.8)$$

where

$$\mathbf{Z}(t_i, \boldsymbol{\theta}) = \frac{\delta_i \mathbf{g}(t_i, \boldsymbol{\theta})}{R_i/n} = (Z_1(t_i, \boldsymbol{\theta}), \dots, Z_r(t_i, \boldsymbol{\theta}))^\top \quad \text{for } i = 1, 2, \dots, n-1, \quad (3.9)$$

and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_r)^\top$ are the solutions of the following constraint equations:

$$\begin{aligned}
\mathbf{h}(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) &= \sum_{i=1}^{n-1} \delta_i \mathbf{g}(t_i, \boldsymbol{\theta}) w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) + \delta_n \mathbf{g}(t_n, \boldsymbol{\theta}) - \mathbf{k} \\
&= \sum_{i=1}^{n-1} \frac{\delta_i \mathbf{g}(t_i, \boldsymbol{\theta})}{R_i} \times \frac{1}{1 + \boldsymbol{\lambda}(\boldsymbol{\theta})^\top \mathbf{Z}(t_i, \boldsymbol{\theta})} + \delta_n \mathbf{g}(t_n, \boldsymbol{\theta}) - \mathbf{k} \\
&= \frac{1}{n} \sum_{i=1}^{n-1} \frac{\mathbf{Z}(t_i, \boldsymbol{\theta})}{1 + \boldsymbol{\lambda}(\boldsymbol{\theta})^\top \mathbf{Z}(t_i, \boldsymbol{\theta})} + \delta_n \mathbf{g}(t_n, \boldsymbol{\theta}) - \mathbf{k} \\
&= \mathbf{0}.
\end{aligned} \tag{3.10}$$

Proof: To use Lagrange multipliers $\boldsymbol{\lambda}$, we form the target function

$$G = \sum_{i=1}^n \delta_i \log w_i - \sum_{i=1}^n w_i R_i - n \boldsymbol{\lambda}^\top \left\{ \sum_{i=1}^{n-1} \mathbf{g}(t_i, \boldsymbol{\theta}) \delta_i w_i + \mathbf{g}(t_n, \boldsymbol{\theta}) \delta_n - \mathbf{k} \right\} \tag{3.11}$$

Taking derivative with respect to w_i and setting to 0 gives

$$\frac{\partial G}{\partial w_i} = \frac{\delta_i}{w_i} - R_i - n \boldsymbol{\lambda}^\top \mathbf{g}(t_i, \boldsymbol{\theta}) \delta_i = 0$$

So,

$$\begin{aligned}
w_i(\boldsymbol{\lambda}, \boldsymbol{\theta}) &= \frac{\delta_i}{R_i + n \boldsymbol{\lambda}^\top \mathbf{g}(t_i, \boldsymbol{\theta}) \delta_i} \\
&= \frac{\delta_i}{R_i} \times \frac{1}{1 + \boldsymbol{\lambda}^\top (\delta_i \mathbf{g}(t_i, \boldsymbol{\theta}) / (R_i/n))} \\
&= \Delta \hat{\Lambda}_{NA}(t_i) \frac{1}{1 + \boldsymbol{\lambda}^\top \mathbf{Z}(t_i, \boldsymbol{\theta})}
\end{aligned}$$

where

$$\mathbf{Z}(t_i, \boldsymbol{\theta}) = \frac{\delta_i \mathbf{g}(t_i, \boldsymbol{\theta})}{R_i/n} = (Z_1(t_i, \boldsymbol{\theta}), \dots, Z_r(t_i, \boldsymbol{\theta}))^\top \quad \text{for } i = 1, 2, \dots, n-1,$$

and $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\theta}) = (\lambda_1, \dots, \lambda_r)^\top$ can be obtained by solving the constraint equations

$$\begin{aligned}
\mathbf{h}(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) &= \sum_{i=1}^{n-1} \delta_i \mathbf{g}(t_i, \boldsymbol{\theta}) w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) + \delta_n \mathbf{g}(t_n, \boldsymbol{\theta}) - \mathbf{k} \\
&= \sum_{i=1}^{n-1} \frac{\delta_i \mathbf{g}(t_i, \boldsymbol{\theta})}{R_i} \times \frac{1}{1 + \boldsymbol{\lambda}(\boldsymbol{\theta})^\top \mathbf{Z}(t_i, \boldsymbol{\theta})} + \delta_n \mathbf{g}(t_n, \boldsymbol{\theta}) - \mathbf{k} \\
&= \frac{1}{n} \sum_{i=1}^{n-1} \frac{\mathbf{Z}(t_i, \boldsymbol{\theta})}{1 + \boldsymbol{\lambda}(\boldsymbol{\theta})^\top \mathbf{Z}(t_i, \boldsymbol{\theta})} + \delta_n \mathbf{g}(t_n, \boldsymbol{\theta}) - \mathbf{k} \\
&= \mathbf{0}. \quad \blacksquare
\end{aligned}$$

Continuing with the setup of section 3.2, we consider the hypothesis:

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \quad \text{vs.} \quad H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0.$$

We propose the empirical-likelihood-ratio test statistic as follows:

$$T = -2 \left\{ \max_{\boldsymbol{\theta} = \boldsymbol{\theta}_0, w_i} \log AL - \max_{\boldsymbol{\theta} \in \mathbb{R}^p, w_i} \log AL \right\}. \quad (3.12)$$

The first maximum in the test statistic above can be obtained through Theorem 1, with the w_i given there and $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. The second maximization above is taken over all possible $\boldsymbol{\theta}$ and w_i . This maximization is more challenging, and we will discuss its computation later.

We state the main theoretical results for the test statistic in Theorems 2 and 3 and some related lemmas below. They are proved mainly through application of Martingale theory and Taylor-expansion. The results in Theorems 2 and 3 are similar to those of Qin and Lawless, but they employ hazard empirical likelihood and they handle right-censored data. Hypothesis tests and confidence intervals can be obtained by using the results. The detailed proofs are deferred to the next section.

Definition: The $\boldsymbol{\theta}$ value that achieves the maximum in the second term of the test statistic T in (3.12) above will be called the maximum-empirical-likelihood estimator, $\hat{\boldsymbol{\theta}}$.

Define the $r \times r$ matrix

$$\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0) = \frac{\partial \mathbf{h}(\boldsymbol{\lambda}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\lambda}} \Big|_{\boldsymbol{\lambda}=\mathbf{0}} = -\frac{1}{n} \sum_{i=1}^{n-1} \mathbf{Z}(t_i, \boldsymbol{\theta}_0) \mathbf{Z}^\top(t_i, \boldsymbol{\theta}_0).$$

It is easy to verify \mathbf{h}' is symmetric and at least non-positive definite.

Lemma 1 *Let $(T_1, \delta_1), \dots, (T_n, \delta_n)$ be n pairs of random variables as defined in (3.3). Suppose $g_u(x, \boldsymbol{\theta})$, $u = 1, \dots, r$, are left-continuous functions and suppose*

$$0 < \int \frac{|g_u(x, \boldsymbol{\theta})| |g_v(x, \boldsymbol{\theta})|}{(1 - F_0(x-))(1 - G_0(x-))} d\Lambda_0(x) < \infty, \quad \forall u, v, \quad 1 \leq u, v \leq r.$$

Then for \mathbf{Z} as defined in (3.9),

$$\frac{1}{n} \sum_{i=1}^n Z_u(t_i, \boldsymbol{\theta}) Z_v(t_i, \boldsymbol{\theta}) = \int \frac{g_u(t, \boldsymbol{\theta}) g_v(t, \boldsymbol{\theta})}{R(t)/n} d\hat{\Lambda}_{NA}(t) \xrightarrow{P} \int \frac{g_u(x, \boldsymbol{\theta}) g_v(x, \boldsymbol{\theta})}{(1 - F_0(x-))(1 - G_0(x-))} d\Lambda_0(x)$$

where

$$R(t) = \sum_{i=1}^n I_{[T_i \geq t]}.$$

Lemma 2 *Under the assumptions of Lemma 1 and under the null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, we have, for \mathbf{Z} defined in (3.9),*

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \mathbf{Z}(t_i, \boldsymbol{\theta}_0) - \mathbf{k}_0 \right) = \sqrt{n} \left(\sum_{i=1}^n \mathbf{g}(t_i, \boldsymbol{\theta}_0) \Delta \hat{\Lambda}_{NA}(t_i) - \mathbf{k}_0 \right) \xrightarrow{\mathcal{D}} MVN(\mathbf{0}, \Sigma_Z),$$

as $n \rightarrow \infty$, where we assume Σ_Z to be positive definite and

$$\mathbf{k}_0 = \left\{ \int g_1(t, \boldsymbol{\theta}_0) d\Lambda_0(t), \dots, \int g_r(t, \boldsymbol{\theta}_0) d\Lambda_0(t) \right\}^\top$$

$$\Sigma_{Zuv} = \int \frac{g_u(x, \boldsymbol{\theta}_0)g_v(x, \boldsymbol{\theta}_0)}{(1 - F_0(x-))(1 - G_0(x-))} d\Lambda_0(x), \quad \forall u, v, \quad 1 \leq u, v \leq r.$$

Lemma 3 *The solution $\boldsymbol{\lambda}$ of the constraint equations in (3.10) under the null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, has the following asymptotic representations:*

(i) *Let $\boldsymbol{\theta}_0$ be the true parameters, and assume that*

$$\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0) = \left. \frac{\partial \mathbf{h}(\boldsymbol{\lambda}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\lambda}} \right|_{\boldsymbol{\lambda}=\mathbf{0}}$$

is an invertible $k \times k$ matrix. Then

$$\sqrt{n}\boldsymbol{\lambda}(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} MVN(\mathbf{0}, \Sigma_\lambda) \quad \text{as } n \rightarrow \infty,$$

where $\Sigma_\lambda = \Sigma_Z^{-1} = \lim_{n \rightarrow \infty} [\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)]^{-1}$.

(ii) *In addition, assume that $\mathbf{g}(\cdot)$ are smooth and $|\boldsymbol{\theta} - \boldsymbol{\theta}_0| = O(1/\sqrt{n})$. Then*

$$\boldsymbol{\lambda}(\boldsymbol{\theta}) = \boldsymbol{\lambda}(\boldsymbol{\theta}_0) - \{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1} A(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_p(1/\sqrt{n})$$

where A is an $r \times p$ matrix defined as

$$A = \left. \frac{\partial \mathbf{h}(\boldsymbol{\lambda}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\lambda}=\mathbf{0}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{1}{n} \sum_{i=1}^n \frac{\partial \mathbf{Z}(t_i, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{R_i/n} \frac{\partial \mathbf{g}(t_i, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$$

Theorem 2 *Under the assumptions of Lemma 1, 2 and 3, the empirical-likelihood-ratio statistic T defined in (3.12) has asymptotically a chi-square distribution with p degrees of freedom under H_0 :*

$$T \xrightarrow{\mathcal{D}} \chi_{(p)}^2, \quad \text{as } n \rightarrow \infty.$$

Theorem 3 Let $\hat{\boldsymbol{\theta}}$ be the maximum empirical likelihood estimator defined above, and $\boldsymbol{\theta}_0$ be the true parameter vector. Suppose $\frac{\partial \mathbf{g}(t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ exists, the rank of $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{R_i/n} \frac{\partial \mathbf{g}(t_i, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$ is p , and $\int \left| \frac{\partial g_i(x, \theta_{0j})}{\partial \theta_j} \right| d\Lambda_0(x) < \infty$, $\forall i = 1, \dots, r$, and $j = 1, \dots, p$. Under the assumptions of Theorem 2, the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ is given by, as $n \rightarrow \infty$,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} MVN(\mathbf{0}, \Sigma_\theta),$$

where

$$\Sigma_\theta = \lim_{n \rightarrow \infty} \{A^\top [-\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)]^{-1} A\}^{-1},$$

$$A = \frac{\partial \mathbf{h}(\boldsymbol{\lambda}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \big|_{\boldsymbol{\lambda}=\mathbf{0}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{1}{n} \sum_{i=1}^n \frac{\partial \mathbf{Z}(t_i, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{R_i/n} \frac{\partial \mathbf{g}(t_i, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$$

We now take a closer look at the asymptotic variance of the maximum-empirical-likelihood estimator obtained in Theorem 3 above. Since the structure of Σ_θ is similar to the structure of the variance-covariance matrix in Qin and Lawless's theorem, we can make the same corollary as they do:

Corollary When $r > p$, the asymptotic variance Σ_θ of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ cannot decrease if a constraint equation is dropped.

Proof: Define

$$\begin{aligned}
D_r^\top(\boldsymbol{\theta}) &= A^\top \\
&= \left(\left(\frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{R_i/n} \frac{\partial g_1(t_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^\top, \dots, \left(\frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{R_i/n} \frac{\partial g_{r-1}(t_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^\top, \left(\frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{R_i/n} \frac{\partial g_r(t_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^\top \right) \\
&= \left(D_{r-1}^\top(\boldsymbol{\theta}), \left(\frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{R_i/n} \frac{\partial g_r(t_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^\top \right),
\end{aligned}$$

$$\begin{aligned}
C_r(\boldsymbol{\theta}) &= -\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}) \\
&= \frac{1}{n} \sum_{i=1}^{n-1} \mathbf{Z}(t_i, \boldsymbol{\theta}) \mathbf{Z}^\top(t_i, \boldsymbol{\theta}) \\
&= \frac{1}{n} \sum_{i=1}^{n-1} \frac{\delta_i}{R_i/n} \mathbf{g}(t_i, \boldsymbol{\theta}) \left(\frac{\delta_i}{R_i/n} \mathbf{g}(t_i, \boldsymbol{\theta}) \right)^\top \\
&= \begin{pmatrix} C_{11}(\boldsymbol{\theta}) & C_{12}(\boldsymbol{\theta}) \\ C_{21}(\boldsymbol{\theta}) & C_{22}(\boldsymbol{\theta}) \end{pmatrix},
\end{aligned}$$

where $C_{11}(\boldsymbol{\theta})$ is an $(r-1) \times (r-1)$ matrix. For square matrices A and B of the same order, let $A \geq B$ denote that $A - B$ is positive semi-definite. Then

$$\begin{aligned}
\Sigma_r^{-1} &= A^\top [-\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta})]^{-1} A \\
&= D_r^\top(\boldsymbol{\theta}) C_r^{-1}(\boldsymbol{\theta}) D_r(\boldsymbol{\theta}) \\
&\geq \left(D_{r-1}^\top(\boldsymbol{\theta}), \left(\frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{R_i/n} \frac{\partial g_r(t_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^\top \right) \begin{pmatrix} C_{11}^{-1}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} D_{r-1}(\boldsymbol{\theta}) \\ \left(\frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{R_i/n} \frac{\partial g_r(t_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \end{pmatrix} \\
&= \Sigma_{r-1}^{-1}. \quad \blacksquare
\end{aligned}$$

This corollary shows that the maximum empirical likelihood estimator $\hat{\boldsymbol{\theta}}$ obtained in Theorem 3 is more efficient (smaller variance-covariance matrix) for the over-determined case than for the just-determined case. The variance-covariance matrix will not decrease when we include fewer constraint equations. In other words, it is recommended that we always use as much information as possible, in the form of estimating equations, at least for large sample sizes.

3.4 Lemmas and Proofs

Lemma 1 *Let $(T_1, \delta_1), \dots, (T_n, \delta_n)$ be n pairs of random variables as defined in (3.3).*

Suppose $g_u(x, \boldsymbol{\theta})$, $u = 1, \dots, r$, are left-continuous functions and suppose

$$0 < \int \frac{|g_u(x, \boldsymbol{\theta})||g_v(x, \boldsymbol{\theta})|}{(1 - F_0(x-))(1 - G_0(x-))} d\Lambda_0(x) < \infty, \quad \forall u, v, \quad 1 \leq u, v \leq r.$$

Then for \mathbf{Z} as defined in (3.9),

$$\frac{1}{n} \sum_{i=1}^n Z_u(t_i, \boldsymbol{\theta}) Z_v(t_i, \boldsymbol{\theta}) = \int \frac{g_u(t, \boldsymbol{\theta}) g_v(t, \boldsymbol{\theta})}{R(t)/n} d\hat{\Lambda}_{NA}(t) \xrightarrow{P} \int \frac{g_u(x, \boldsymbol{\theta}) g_v(x, \boldsymbol{\theta})}{(1 - F_0(x-))(1 - G_0(x-))} d\Lambda_0(x)$$

where

$$R(t) = \sum_{i=1}^n I_{[T_i \geq t]}.$$

Proof: From (3.9),

$$Z_u(t_i, \boldsymbol{\theta}) = \frac{\delta_i g_u(t_i, \boldsymbol{\theta})}{R_i/n}.$$

Therefore,

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n Z_u(t_i, \boldsymbol{\theta}) Z_v(t_i, \boldsymbol{\theta}) &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i g_u(t_i, \boldsymbol{\theta})}{R_i/n} \times \frac{\delta_i g_v(t_i, \boldsymbol{\theta})}{R_i/n} \\
&= \sum_{i=1}^n \frac{g_u(t_i, \boldsymbol{\theta}) g_v(t_i, \boldsymbol{\theta})}{R_i/n} \frac{\delta_i}{R_i} \\
&= \sum_{i=1}^n \frac{g_u(t_i, \boldsymbol{\theta}) g_v(t_i, \boldsymbol{\theta})}{R_i/n} \Delta \hat{\Lambda}_{NA}(t_i) \\
&= \int \frac{g_u(t, \boldsymbol{\theta}) g_v(t, \boldsymbol{\theta})}{R(t)/n} d\hat{\Lambda}_{NA}(t).
\end{aligned}$$

And from Lemma 3.7 in Pan (1997) [27],

$$\sum_{i=1}^n \frac{g_u(t_i, \boldsymbol{\theta}) g_v(t_i, \boldsymbol{\theta})}{R_i/n} \Delta \hat{\Lambda}_{NA}(t_i) \xrightarrow{P} \int \frac{g_u(x, \boldsymbol{\theta}) g_v(x, \boldsymbol{\theta})}{(1 - F_0(x-))(1 - G_0(x-))} d\Lambda_0(x),$$

which completes the proof. \blacksquare

Lemma 2 *Under the assumptions of Lemma 1 and under the null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, we have, for \mathbf{Z} defined in (3.9),*

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \mathbf{Z}(t_i, \boldsymbol{\theta}_0) - \mathbf{k}_0 \right) = \sqrt{n} \left(\sum_{i=1}^n \mathbf{g}(t_i, \boldsymbol{\theta}_0) \Delta \hat{\Lambda}_{NA}(t_i) - \mathbf{k}_0 \right) \xrightarrow{\mathcal{D}} MVN(\mathbf{0}, \Sigma_{\mathbf{Z}}),$$

as $n \rightarrow \infty$, where

$$\begin{aligned}
\mathbf{k}_0 &= \left\{ \int g_1(t, \boldsymbol{\theta}_0) d\Lambda_0(t), \dots, \int g_r(t, \boldsymbol{\theta}_0) d\Lambda_0(t) \right\}^\top \\
\Sigma_{Z_{uv}} &= \int \frac{g_u(x, \boldsymbol{\theta}_0) g_v(x, \boldsymbol{\theta}_0)}{(1 - F_0(x-))(1 - G_0(x-))} d\Lambda_0(x), \quad \forall u, v, \quad 1 \leq u, v \leq r.
\end{aligned}$$

Proof: We know from (3.9), for $1 \leq u \leq r$,

$$\begin{aligned}
\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n Z_u(t_i, \boldsymbol{\theta}_0) - k_{0u} \right) &= \sqrt{n} \left(\sum_{i=1}^n g_u(t_i, \boldsymbol{\theta}_0) \Delta \hat{\Lambda}_{NA}(t_i) - k_{0u} \right) \\
&= \sqrt{n} \left(\int g_u(t, \boldsymbol{\theta}_0) d\hat{\Lambda}_{NA}(t) - \int g_u(t, \boldsymbol{\theta}_0) d\Lambda_0(t) \right).
\end{aligned}$$

By Lemma 3.3 in Pan (1997), the Lindeberg's condition holds for each Z_u , $u = 1, \dots, r$, and

$$\sqrt{n} \left(\int g_u(t, \boldsymbol{\theta}_0) d\hat{\Lambda}_{NA}(t) - \int g_u(t, \boldsymbol{\theta}_0) d\Lambda_0(t) \right) \xrightarrow{\mathcal{D}} N(0, \sigma_u^2)$$

as $n \rightarrow \infty$, where

$$\sigma_u^2 = \int \frac{g_u^2(x, \boldsymbol{\theta}_0)}{(1 - F_0(x-))(1 - G_0(x-))} d\Lambda_0(x).$$

By (4.14) in Anderson et. al. (1993) [2] p.178,

$$\hat{\Lambda}_{NA}(t) - \Lambda_0(t) = \int_0^t \frac{J(s)}{R(s)} dM(s),$$

where $J(s) = I_{[R(s)>0]}$ and $M(s)$ is a local square integrable martingale endowed with a filter \mathcal{F}_t .

Therefore,

$$\begin{aligned} \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n Z_u(t_i, \boldsymbol{\theta}_0) - k_{0u} \right) &= \sqrt{n} \int g_u(t, \boldsymbol{\theta}_0) d \left(\hat{\Lambda}_{NA}(t) - \Lambda_0(t) \right) \\ &= \sqrt{n} \int g_u(t, \boldsymbol{\theta}_0) \frac{J(t)}{R(t)} dM(t), \end{aligned}$$

and the predictable covariance process of $\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n Z_u(t_i, \boldsymbol{\theta}_0) - k_{0u} \right)$ and $\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n Z_v(t_i, \boldsymbol{\theta}_0) - k_{0v} \right) \quad \forall u \neq v$ is

$$\int g_u(t, \boldsymbol{\theta}_0) g_v(t, \boldsymbol{\theta}_0) d \left\langle \sqrt{n} \int_0^\infty \frac{J(t)}{R(t)} dM(t) \right\rangle.$$

Observe that

$$\begin{aligned} \left\langle \sqrt{n} \int_0^\infty \frac{J(t)}{R(t)} dM(t) \right\rangle &= \int_0^\infty \frac{J^2(t)}{R(t)/n} d\Lambda_0 \\ &\xrightarrow{\mathcal{P}} \int_0^\infty \frac{d\Lambda_0(x)}{(1 - F_0(x))(1 - G_0(x))}, \end{aligned}$$

hence,

$$\begin{aligned}
& \left\langle \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n Z_u(t_i, \boldsymbol{\theta}_0) - k_{0u} \right), \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n Z_v(t_i, \boldsymbol{\theta}_0) - k_{0v} \right) \right\rangle \\
&= \int \frac{g_u(x, \boldsymbol{\theta}_0) g_v(x, \boldsymbol{\theta}_0) J^2(t)}{R(t)/n} d\Lambda_0 \\
&\rightarrow \int \frac{g_u(x, \boldsymbol{\theta}_0) g_v(x, \boldsymbol{\theta}_0)}{(1 - F_0(x-))(1 - G_0(x-))} d\Lambda_0(x).
\end{aligned}$$

Therefore, by Multivariate Counting Process Martingale Central Limit Theorem,

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \mathbf{Z}(t_i, \boldsymbol{\theta}_0) - \mathbf{k}_0 \right) \xrightarrow{\mathcal{D}} MVN(\mathbf{0}, \Sigma_Z)$$

where

$$\begin{aligned}
\mathbf{k}_0 &= \left\{ \int g_1(t, \boldsymbol{\theta}_0) d\Lambda_0(t), \dots, \int g_r(t, \boldsymbol{\theta}_0) d\Lambda_0(t) \right\}^\top \\
\Sigma_{Zuv} &= \int \frac{g_u(x, \boldsymbol{\theta}_0) g_v(x, \boldsymbol{\theta}_0)}{(1 - F_0(x-))(1 - G_0(x-))} d\Lambda_0(x) \quad \forall u, v \quad 1 \leq u, v \leq r. \quad \blacksquare
\end{aligned}$$

Lemma 3 *The solution $\boldsymbol{\lambda}$ of the constraint equations in (3.10) under the null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, has the following asymptotic representations:*

(i) *Let $\boldsymbol{\theta}_0$ be the true parameters, and assume that*

$$\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0) = \frac{\partial \mathbf{h}(\boldsymbol{\lambda}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\lambda}} \Big|_{\boldsymbol{\lambda}=\mathbf{0}}$$

is an invertible $k \times k$ matrix. Then

$$\sqrt{n} \boldsymbol{\lambda}(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} MVN(\mathbf{0}, \Sigma_\lambda) \quad \text{as } n \rightarrow \infty,$$

where $\Sigma_\lambda = \Sigma_Z^{-1} = \lim_{n \rightarrow \infty} [\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)]^{-1}$.

(ii) *In addition, assume that $\mathbf{g}(\cdot)$ are smooth and $|\boldsymbol{\theta} - \boldsymbol{\theta}_0| = O(1/\sqrt{n})$. Then*

$$\boldsymbol{\lambda}(\boldsymbol{\theta}) = \boldsymbol{\lambda}(\boldsymbol{\theta}_0) - \{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1} A(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_p(1/\sqrt{n})$$

where A is an $r \times p$ matrix defined as

$$A = \frac{\partial \mathbf{h}(\boldsymbol{\lambda}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\lambda}=\mathbf{0}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{1}{n} \sum_{i=1}^n \frac{\partial \mathbf{Z}(t_i, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{R_i/n} \frac{\partial \mathbf{g}(t_i, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$$

Proof: Define

$$\mathbf{h}(\boldsymbol{\lambda}(\mathbf{s}), \mathbf{s}) = \sum_{i=1}^{n-1} \delta_i \mathbf{g}(t_i, \mathbf{s}) w_i(\boldsymbol{\lambda}(\mathbf{s}), \mathbf{s}) + \delta_n \mathbf{g}(t_n, \mathbf{s}) - \mathbf{k}.$$

Under the null hypothesis, $\mathbf{h}(\boldsymbol{\lambda}(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) = \mathbf{0}$. Using a Taylor expansion of the first variable at $\mathbf{0}$, we can write

$$\begin{aligned} \mathbf{0} &= \mathbf{h}(\boldsymbol{\lambda}(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) \\ &= \mathbf{h}(\mathbf{0}, \boldsymbol{\theta}_0) + \mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)(\boldsymbol{\lambda}(\boldsymbol{\theta}_0) - \mathbf{0}) + o_p(1/\sqrt{n}), \end{aligned}$$

where $\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)$ is an $r \times r$ invertible matrix. Rearranging the above equation, we obtain

$$\sqrt{n}\boldsymbol{\lambda}(\boldsymbol{\theta}_0) = -\{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1}(\sqrt{n}\mathbf{h}(\mathbf{0}, \boldsymbol{\theta}_0)) + o_p(1),$$

where

$$\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0) = \frac{\partial \mathbf{h}(\boldsymbol{\lambda}(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)}{\partial \boldsymbol{\lambda}(\boldsymbol{\theta}_0)} \Big|_{\boldsymbol{\lambda}(\boldsymbol{\theta}_0)=\mathbf{0}} = -\frac{1}{n} \sum_{i=1}^{n-1} \mathbf{Z}(t_i, \boldsymbol{\theta}_0) \mathbf{Z}^\top(t_i, \boldsymbol{\theta}_0).$$

The elements of $\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)$ are explicitly computed as

$$\begin{aligned} h'_{uv}(\mathbf{0}, \boldsymbol{\theta}_0) &= -\sum_{i=1}^{n-1} \frac{ng_u(t_i, \boldsymbol{\theta}_0)g_v(t_i, \boldsymbol{\theta}_0)\delta_i}{R_i^2} \\ &= -\frac{1}{n} \sum_{i=1}^{n-1} Z_u(t_i, \boldsymbol{\theta}_0)Z_v(t_i, \boldsymbol{\theta}_0). \end{aligned}$$

Then by Lemma 1 and Lemma 2, we have as $n \rightarrow \infty$,

$$\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0) \xrightarrow{P} -\Sigma_Z$$

Since

$$\begin{aligned}\mathbf{h}(\mathbf{0}, \boldsymbol{\theta}_0) &= \sum_{i=1}^{n-1} \frac{\delta_i \mathbf{g}(t_i, \boldsymbol{\theta}_0)}{R_i} + \delta_n \mathbf{g}(t_n, \boldsymbol{\theta}_0) - \mathbf{k} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{Z}(t_i, \boldsymbol{\theta}_0) - \mathbf{k},\end{aligned}$$

by Lemma 2, we have

$$\sqrt{n} \mathbf{h}(\mathbf{0}, \boldsymbol{\theta}_0) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \mathbf{Z}(t_i, \boldsymbol{\theta}_0) - \mathbf{k} \right) \xrightarrow{\mathcal{D}} MVN(\mathbf{0}, \Sigma_Z).$$

So

$$\sqrt{n} \boldsymbol{\lambda}(\boldsymbol{\theta}_0) = -\{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1} (\sqrt{n} \mathbf{h}(\mathbf{0}, \boldsymbol{\theta}_0)) + o_p(1) \xrightarrow{\mathcal{D}} MVN(\mathbf{0}, \Sigma_\lambda)$$

with

$$\Sigma_\lambda = -[\lim \mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)]^{-1} = \Sigma_Z^{-1}$$

Use the same vector function $\mathbf{h}(\boldsymbol{\lambda}(\mathbf{s}), \mathbf{s})$. We have $\mathbf{h}(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) = \mathbf{0}$ due to the constraint we have. Using a Taylor expansion at $(\mathbf{0}, \boldsymbol{\theta}_0)$ we can write

$$\begin{aligned}\mathbf{0} &= \mathbf{h}(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) \\ &= \mathbf{h}(\mathbf{0}, \boldsymbol{\theta}_0) + [\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0), A](\boldsymbol{\lambda}(\boldsymbol{\theta}) - \mathbf{0}, \boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top + o_p(1/\sqrt{n}),\end{aligned}$$

where

$$A = \frac{\partial \mathbf{h}(\boldsymbol{\lambda}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\lambda}=\mathbf{0}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{1}{n} \sum_{i=1}^n \frac{\partial \mathbf{Z}(t_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.$$

Setting $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ in the above gives

$$\begin{aligned}\mathbf{0} &= \mathbf{h}(\boldsymbol{\lambda}(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) \\ &= \mathbf{h}(\mathbf{0}, \boldsymbol{\theta}_0) + \mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)(\boldsymbol{\lambda}(\boldsymbol{\theta}_0) - \mathbf{0})^\top + o_p(1/\sqrt{n}).\end{aligned}$$

Combining the two equations above gives

$$\begin{aligned}\mathbf{0} &= \mathbf{h}(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) - \mathbf{h}(\boldsymbol{\lambda}(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) \\ &= \mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)[\boldsymbol{\lambda}(\boldsymbol{\theta}) - \boldsymbol{\lambda}(\boldsymbol{\theta}_0)] + A(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_p(1/\sqrt{n})\end{aligned}$$

and therefore,

$$\boldsymbol{\lambda}(\boldsymbol{\theta}) - \boldsymbol{\lambda}(\boldsymbol{\theta}_0) = -\{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1}A(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_p(1/\sqrt{n}).$$

Hence,

$$\boldsymbol{\lambda}(\boldsymbol{\theta}) = \boldsymbol{\lambda}(\boldsymbol{\theta}_0) - \{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1}A(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_p(1/\sqrt{n}). \quad \blacksquare$$

Proof of Theorems 2 and 3

Proof: Let us define

$$f(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) = \sum_{i=1}^n \delta_i \log w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) - \sum_{i=1}^n R_i w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}).$$

The log empirical-likelihood-ratio test statistic takes the form

$$T = -2 \left\{ f(\boldsymbol{\lambda}(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) - \max_{\boldsymbol{\theta}} f(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) \right\}.$$

Using a Taylor expansion we obtain

$$\begin{aligned}
T &= -2\{f(\mathbf{0}, \boldsymbol{\theta}_0) \\
&+ (\boldsymbol{\lambda}(\boldsymbol{\theta}_0), \mathbf{0}) \left(\frac{\partial f(\boldsymbol{\lambda}(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)}{\partial \boldsymbol{\lambda}(\boldsymbol{\theta}_0)}, \frac{\partial f(\boldsymbol{\lambda}(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0} \right)^\top \Big|_{\boldsymbol{\lambda}(\boldsymbol{\theta}_0)=\mathbf{0}} \\
&+ \frac{1}{2} (\boldsymbol{\lambda}(\boldsymbol{\theta}_0), \mathbf{0}) D (\boldsymbol{\lambda}(\boldsymbol{\theta}_0), \mathbf{0})^\top + o_p(1) \\
&- \max_{\boldsymbol{\theta}} \{f(\mathbf{0}, \boldsymbol{\theta}_0) \\
&+ (\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta} - \boldsymbol{\theta}_0) \left(\frac{\partial f(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\lambda}(\boldsymbol{\theta})}, \frac{\partial f(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^\top \Big|_{\boldsymbol{\lambda}(\boldsymbol{\theta})=\mathbf{0}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&+ \frac{1}{2} (\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta} - \boldsymbol{\theta}_0) D (\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top + o_p(1)\} \}
\end{aligned}$$

where D denotes the $(r+p) \times (r+p)$ matrix of the second derivatives of $f(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})$ with respect to $\boldsymbol{\lambda}(\boldsymbol{\theta})$ and $\boldsymbol{\theta}$ as follows:

$$D = \begin{pmatrix} \frac{\partial^2}{\partial \boldsymbol{\lambda}(\boldsymbol{\theta})^2} f(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) \Big|_{\boldsymbol{\lambda}(\boldsymbol{\theta})=\mathbf{0}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} & \frac{\partial^2}{\partial \boldsymbol{\lambda}(\boldsymbol{\theta}) \partial \boldsymbol{\theta}} f(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) \Big|_{\boldsymbol{\lambda}(\boldsymbol{\theta})=\mathbf{0}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ \frac{\partial^2}{\partial \boldsymbol{\lambda}(\boldsymbol{\theta}) \partial \boldsymbol{\theta}} f(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) \Big|_{\boldsymbol{\lambda}(\boldsymbol{\theta})=\mathbf{0}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} & \frac{\partial^2}{\partial \boldsymbol{\theta}^2} f(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) \Big|_{\boldsymbol{\lambda}(\boldsymbol{\theta})=\mathbf{0}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} \end{pmatrix}$$

We will calculate all the derivatives below. Notice that when $\boldsymbol{\lambda}(\boldsymbol{\theta}) = \mathbf{0}$, $w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) = \frac{\delta_i}{R_i}$. The first derivatives for D are calculated as follows:

$$\begin{aligned}
\frac{\partial f(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\lambda}(\boldsymbol{\theta})} \Big|_{\boldsymbol{\lambda}(\boldsymbol{\theta})=\mathbf{0}} &= \sum_{i=1}^n \frac{\delta_i}{w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})} \cdot \frac{\partial w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} - \sum_{i=1}^n R_i \cdot \frac{\partial w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} \Big|_{\boldsymbol{\lambda}(\boldsymbol{\theta})=\mathbf{0}} \\
&= \sum_{i=1}^n R_i \cdot \frac{\partial w_i(\mathbf{0}, \boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} - \sum_{i=1}^n R_i \cdot \frac{\partial w_i(\mathbf{0}, \boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} \\
&= \mathbf{0}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial f(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\lambda}(\boldsymbol{\theta})=\mathbf{0}} &= \sum_{i=1}^n \frac{\delta_i}{w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})} \cdot \frac{\partial w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \sum_{i=1}^n R_i \cdot \frac{\partial w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\lambda}(\boldsymbol{\theta})=\mathbf{0}} \\
&= \sum_{i=1}^n R_i \cdot \frac{\partial w_i(\mathbf{0}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \sum_{i=1}^n R_i \cdot \frac{\partial w_i(\mathbf{0}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\
&= \mathbf{0}
\end{aligned}$$

The second derivatives for D are calculated as follows:

$$\begin{aligned}
&\frac{\partial^2}{\partial \boldsymbol{\lambda}(\boldsymbol{\theta})^2} f(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) \Big|_{\boldsymbol{\lambda}(\boldsymbol{\theta})=\mathbf{0}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= \sum_{i=1}^n -\frac{\delta_i}{w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})^2} \cdot \left(\frac{\partial w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} \right) \left(\frac{\partial w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} \right)^\top + \sum_{i=1}^n \frac{\delta_i}{w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})} \cdot \frac{\partial^2 w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\lambda}^2} \\
&\quad - \sum_{i=1}^n R_i \cdot \frac{\partial^2 w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\lambda}^2} \Big|_{\boldsymbol{\lambda}(\boldsymbol{\theta})=\mathbf{0}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= -\sum_{i=1}^n \frac{R_i^2}{\delta_i} \cdot \frac{n^2 \delta_i^2 \mathbf{g}(t_i, \boldsymbol{\theta}_0) \mathbf{g}(t_i, \boldsymbol{\theta}_0)^\top}{R_i^4} + \sum_{i=1}^n R_i \cdot \frac{\partial^2 w_i(\mathbf{0}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\lambda}^2} - \sum_{i=1}^n R_i \cdot \frac{\partial^2 w_i(\mathbf{0}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\lambda}^2} \\
&= -\sum_{i=1}^n \frac{R_i^2}{\delta_i} \cdot \frac{n^2 \delta_i^2 \mathbf{g}(t_i, \boldsymbol{\theta}_0) \mathbf{g}(t_i, \boldsymbol{\theta}_0)^\top}{R_i^4} \\
&= -\sum_{i=1}^n \frac{\delta_i \mathbf{g}(t_i, \boldsymbol{\theta}_0)}{R_i/n} \left(\frac{\delta_i \mathbf{g}(t_i, \boldsymbol{\theta}_0)}{R_i/n} \right)^\top \\
&= -\sum_{i=1}^n \mathbf{Z}(t_i, \boldsymbol{\theta}_0) (\mathbf{Z}(t_i, \boldsymbol{\theta}_0))^\top \\
&= n \mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)
\end{aligned}$$

$$\begin{aligned}
& \frac{\partial^2}{\partial \boldsymbol{\lambda}(\boldsymbol{\theta}) \partial \boldsymbol{\theta}} f(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) \Big|_{\boldsymbol{\lambda}(\boldsymbol{\theta})=\mathbf{0}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= - \sum_{i=1}^n \frac{\delta_i}{w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})^2} \cdot \frac{\partial w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \cdot \frac{\partial w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} \\
&+ \sum_{i=1}^n \frac{\delta_i}{w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})} \cdot \frac{\partial^2 w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\theta}} - \sum_{i=1}^n R_i \cdot \frac{\partial^2 w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\lambda}(\boldsymbol{\theta})=\mathbf{0}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= \mathbf{0},
\end{aligned}$$

since

$$\frac{\partial w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\lambda}(\boldsymbol{\theta})=\mathbf{0}} = - \frac{n \boldsymbol{\lambda}(\boldsymbol{\theta})^\top \cdot \frac{\partial \mathbf{g}(t_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}}{(R_i + n \boldsymbol{\lambda}(\boldsymbol{\theta})^\top \mathbf{g}(t_i, \boldsymbol{\theta}) \delta_i)^2} \Big|_{\boldsymbol{\lambda}(\boldsymbol{\theta})=\mathbf{0}} = \mathbf{0}.$$

Similarly,

$$\begin{aligned}
& \frac{\partial^2}{\partial \boldsymbol{\theta}^2} f(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) \Big|_{\boldsymbol{\lambda}(\boldsymbol{\theta})=\mathbf{0}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= \sum_{i=1}^n - \frac{\delta_i}{w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})^2} \cdot \left(\frac{\partial w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^\top \\
&+ \sum_{i=1}^n \frac{\delta_i}{w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})} \cdot \frac{\partial^2 w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} - \sum_{i=1}^n R_i \cdot \frac{\partial^2 w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \Big|_{\boldsymbol{\lambda}(\boldsymbol{\theta})=\mathbf{0}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= \mathbf{0}
\end{aligned}$$

Hence,

$$D = \begin{pmatrix} n \mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

And therefore,

$$\begin{aligned} T &= -\{\boldsymbol{\lambda}(\boldsymbol{\theta}_0)^\top n\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\boldsymbol{\lambda}(\boldsymbol{\theta}_0) + o_p(1) - \max_{\boldsymbol{\theta}}\{\boldsymbol{\lambda}(\boldsymbol{\theta})^\top n\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\boldsymbol{\lambda}(\boldsymbol{\theta})\} + o_p(1)\} \\ &= \max_{\boldsymbol{\theta}}\{\boldsymbol{\lambda}(\boldsymbol{\theta})^\top n\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\boldsymbol{\lambda}(\boldsymbol{\theta})\} - \boldsymbol{\lambda}(\boldsymbol{\theta}_0)^\top n\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\boldsymbol{\lambda}(\boldsymbol{\theta}_0) + o_p(1) \end{aligned}$$

Since by Lemma 3,

$$\boldsymbol{\lambda}(\boldsymbol{\theta}) = \boldsymbol{\lambda}(\boldsymbol{\theta}_0) - \{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1}A(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_p(1/\sqrt{n}) ,$$

therefore we can write

$$\begin{aligned} T &= \max_{\boldsymbol{\theta}}\{(\boldsymbol{\lambda}(\boldsymbol{\theta}_0) - \{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1}A(\boldsymbol{\theta} - \boldsymbol{\theta}_0))^\top n\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)(\boldsymbol{\lambda}(\boldsymbol{\theta}_0) - \{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1}A(\boldsymbol{\theta} - \boldsymbol{\theta}_0))\} \\ &\quad - \boldsymbol{\lambda}(\boldsymbol{\theta}_0)^\top n\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\boldsymbol{\lambda}(\boldsymbol{\theta}_0) + o_p(1). \end{aligned}$$

Maximization over $\boldsymbol{\theta}$ can be accomplished via the following lemma. It can easily be proved by linear algebra.

Lemma 4 *Suppose Γ is a negative definite $r \times r$ matrix and q is an $r \times 1$ vector. Also suppose B is an $r \times p$ matrix and u is a $p \times 1$ vector.*

The quadratic form

$$\max_u (q - Bu)^\top \Gamma (q - Bu)$$

is maximized when $u = u^$ is the solution of the equation $B^\top \Gamma (q - Bu) = 0$, which is equivalent to $u^* = (B^\top \Gamma B)^{-1} B^\top \Gamma q$. And the maximum value achieved is*

$$q^\top \Gamma q - (Bu^*)^\top \Gamma (Bu^*).$$

In this case,

$$q = \boldsymbol{\lambda}(\boldsymbol{\theta}_0)$$

$$B = \{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1}A$$

$$u = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)$$

$$\Gamma = n\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0),$$

then T becomes

$$\max_u (q - Bu)^\top \Gamma (q - Bu) - q^\top \Gamma q + o_p(1)$$

Since by Lemma 4 the maximization of the quadratic form occurs when $u = u^*$ is the solution of the equation $B^\top \Gamma (q - Bu) = 0$ and $u^* = (B^\top \Gamma B)^{-1} B^\top \Gamma q$, and also notice that $\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)$ is symmetric, so the maximization of T occurs when

$$\begin{aligned} \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 &= u^* \\ &= (B^\top \Gamma B)^{-1} B^\top \Gamma q \\ &= \left\{ A^\top \{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1} n \mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0) \{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1} A \right\}^{-1} \\ &\quad A^\top \{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1} n \{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0) \boldsymbol{\lambda}(\boldsymbol{\theta}_0)\} \\ &= \left\{ A^\top \{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1} A \right\}^{-1} A^\top \boldsymbol{\lambda}(\boldsymbol{\theta}_0) \end{aligned}$$

And the maximum value of T is

$$\begin{aligned}
T^* &= q^\top \Gamma q - (Bu^*)^\top \Gamma (Bu^*) - q^\top \Gamma q + o_p(1) \\
&= -(Bu^*)^\top \Gamma (Bu^*) + o_p(1) \\
&= -u^{*\top} A^\top \{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1} n \mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0) \{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1} Au^* + o_p(1) \\
&= -n(\boldsymbol{\lambda}(\boldsymbol{\theta}_0))^\top A \left\{ A^\top \{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1} A \right\}^{-1} A^\top \{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1} A \\
&\quad \left\{ A^\top \{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1} A \right\}^{-1} A^\top \boldsymbol{\lambda}(\boldsymbol{\theta}_0) + o_p(1) \\
&= (\sqrt{n} \boldsymbol{\lambda}(\boldsymbol{\theta}_0))^\top A \left\{ A^\top \{-\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)^{-1}\} A \right\}^{-1} A^\top (\sqrt{n} \boldsymbol{\lambda}(\boldsymbol{\theta}_0)) + o_p(1),
\end{aligned}$$

where

$$A = \frac{\partial \mathbf{h}(\boldsymbol{\lambda}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\lambda}=\mathbf{0}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{1}{n} \sum_{i=1}^n \frac{\partial \mathbf{Z}(t_i, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}},$$

and

$$\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0) = \frac{\partial \mathbf{h}(\boldsymbol{\lambda}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\lambda}} \Big|_{\boldsymbol{\lambda}=\mathbf{0}} = -\frac{1}{n} \sum_{i=1}^{n-1} \mathbf{Z}(t_i, \boldsymbol{\theta}_0) \mathbf{Z}^\top(t_i, \boldsymbol{\theta}_0)$$

is an $r \times r$ non-positive matrix.

We can finish the proof using Lemma 3 along with Lemma 5 below.

Lemma 5 (Graybill 1976 [9]) *Suppose $Y \xrightarrow{\mathcal{D}} MVN(0, V)$ and M is a symmetric matrix. Then $YMY^\top \xrightarrow{\mathcal{D}} \chi_p^2$ if and only if MV is idempotent and $\text{rank}(MV) = p$.*

By Lemma 3, we know

$$Y = \sqrt{n}\boldsymbol{\lambda}(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} MVN(\mathbf{0}, \Sigma_\lambda) \quad \text{as } n \rightarrow \infty$$

where

$$V = \Sigma_\lambda = \Sigma_Z^{-1} = \lim_{n \rightarrow \infty} \{-\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1}.$$

Also let

$$M = \lim_{n \rightarrow \infty} A \left\{ A^\top \{-\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)^{-1}\} A \right\}^{-1} A^\top.$$

Since

$$\begin{aligned} MV MV &= \lim_{n \rightarrow \infty} A \left\{ A^\top \{-\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)^{-1}\} A \right\}^{-1} A^\top \lim_{n \rightarrow \infty} \{-\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1} \\ &\quad \lim_{n \rightarrow \infty} A \left\{ A^\top \{-\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)^{-1}\} A \right\}^{-1} A^\top \lim_{n \rightarrow \infty} \{-\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1} \\ &= \lim_{n \rightarrow \infty} A \left\{ A^\top \{-\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)^{-1}\} A \right\}^{-1} A^\top \lim_{n \rightarrow \infty} \{-\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1} \\ &= MV. \end{aligned}$$

so MV is idempotent. Since $g_i(t, \cdot), i = 1, \dots, r$ are functionally independent functions, so the rank of $\lim_{n \rightarrow \infty} \{-\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}$ is r . Since the rank of $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{R_i/n} \frac{\partial \mathbf{g}(t_i, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$ = $\lim_{n \rightarrow \infty} A$ is p , and $p < r$, so $\text{rank}(MV) = p$. Hence,

$$T = YMY^\top \xrightarrow{\mathcal{D}} \chi_p^2.$$

Since

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \left\{ A^\top \{\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1} A \right\}^{-1} A^\top \boldsymbol{\lambda}(\boldsymbol{\theta}_0),$$

and by the asymptotic normality result on $\boldsymbol{\lambda}(\boldsymbol{\theta}_0)$ in Lemma 3, a straightforward calculation shows that

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} MVN(\mathbf{0}, \Sigma_\theta) ,$$

where

$$\begin{aligned} \Sigma_\theta &= \lim_{n \rightarrow \infty} \{A^\top [-\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0)]^{-1} A\}^{-1}, \\ A &= \frac{\partial \mathbf{h}(\boldsymbol{\lambda}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\lambda}=\mathbf{0}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{1}{n} \sum_{i=1}^n \frac{\partial \mathbf{Z}(t_i, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{R_i/n} \frac{\partial \mathbf{g}(t_i, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}. \quad \blacksquare \end{aligned}$$

Finally we give a proof that $\|\boldsymbol{\lambda}(\boldsymbol{\theta}_0)\| = O_p(n^{-1/2})$, similar to the proof in Owen (1990) [23], which will validate the Taylor expansions above. Pan and Zhou (2002) [29] proved that for each Z_k , $k = 1, \dots, r$, $\max_{1 \leq i \leq n} Z_k(t_i, \boldsymbol{\theta}_0) = o_p(n^{1/2})$, and since r is a fixed positive integer and $r \ll n$, so

$$\max_{1 \leq i \leq n} \|\mathbf{Z}(t_i, \boldsymbol{\theta}_0)\| = o_p(n^{1/2}) \tag{3.13}$$

Lemma 6 *With the same notations as above, we have $\|\boldsymbol{\lambda}(\boldsymbol{\theta}_0)\| = O_p(n^{-1/2})$.*

Proof: Let $\boldsymbol{\lambda}(\boldsymbol{\theta}_0) = \rho \boldsymbol{\phi}$ where $\rho \geq 0$ and $\|\boldsymbol{\phi}\| = 1$. Then,

$$\begin{aligned}
\mathbf{0} &= \|\mathbf{h}(\boldsymbol{\lambda}(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)\| \\
&= \left\| \frac{1}{n} \sum_{i=1}^{n-1} \frac{\mathbf{Z}(t_i, \boldsymbol{\theta}_0)}{1 + \boldsymbol{\lambda}(\boldsymbol{\theta}_0)^\top \mathbf{Z}(t_i, \boldsymbol{\theta}_0)} + \delta_n \mathbf{g}(t_n, \boldsymbol{\theta}_0) - \mathbf{k} \right\| \\
&= \left\| \frac{1}{n} \sum_{i=1}^{n-1} \frac{\mathbf{Z}(t_i, \boldsymbol{\theta}_0)}{1 + \boldsymbol{\lambda}(\boldsymbol{\theta}_0)^\top \mathbf{Z}(t_i, \boldsymbol{\theta}_0)} + \frac{1}{n} \mathbf{Z}(t_n, \boldsymbol{\theta}_0) - \mathbf{k} \right\| \\
&= \left\| \frac{1}{n} \sum_{i=1}^{n-1} \mathbf{Z}(t_i, \boldsymbol{\theta}_0) - \frac{\boldsymbol{\lambda}^\top(\boldsymbol{\theta}_0)}{n} \sum_{i=1}^{n-1} \frac{\mathbf{Z}(t_i, \boldsymbol{\theta}_0) \mathbf{Z}^\top(t_i, \boldsymbol{\theta}_0)}{1 + \boldsymbol{\lambda}(\boldsymbol{\theta}_0)^\top \mathbf{Z}(t_i, \boldsymbol{\theta}_0)} + \frac{1}{n} \mathbf{Z}(t_n, \boldsymbol{\theta}_0) - \mathbf{k} \right\| \\
&= \left\| \left(\frac{1}{n} \sum_{i=1}^n \mathbf{Z}(t_i, \boldsymbol{\theta}_0) - \mathbf{k} \right) - \frac{\boldsymbol{\lambda}^\top(\boldsymbol{\theta}_0)}{n} \sum_{i=1}^{n-1} \frac{\mathbf{Z}(t_i, \boldsymbol{\theta}_0) \mathbf{Z}^\top(t_i, \boldsymbol{\theta}_0)}{1 + \boldsymbol{\lambda}(\boldsymbol{\theta}_0)^\top \mathbf{Z}(t_i, \boldsymbol{\theta}_0)} \right\| \\
&\geq \frac{1}{n} \left| \boldsymbol{\phi}^\top \left(\left(\sum_{i=1}^n \mathbf{Z}(t_i, \boldsymbol{\theta}_0) - \mathbf{k} \right) - \rho \sum_{i=1}^{n-1} \frac{\mathbf{Z}(t_i, \boldsymbol{\theta}_0) \boldsymbol{\phi}^\top \mathbf{Z}(t_i, \boldsymbol{\theta}_0)}{1 + \rho \boldsymbol{\phi}^\top \mathbf{Z}(t_i, \boldsymbol{\theta}_0)} \right) \right| \\
&\geq \frac{\rho}{n} \boldsymbol{\phi}^\top \sum_{i=1}^{n-1} \frac{\mathbf{Z}(t_i, \boldsymbol{\theta}_0) \mathbf{Z}^\top(t_i, \boldsymbol{\theta}_0)}{1 + \rho \boldsymbol{\phi}^\top \mathbf{Z}(t_i, \boldsymbol{\theta}_0)} \boldsymbol{\phi} - \frac{1}{n} \left| \sum_{j=1}^r \mathbf{e}_j^\top \left(\sum_{i=1}^n \mathbf{Z}(t_i, \boldsymbol{\theta}_0) - \mathbf{k} \right) \right| \\
&\geq \frac{\rho \boldsymbol{\phi}^\top S \boldsymbol{\phi}}{1 + \rho B_n} - \frac{1}{n} \left| \sum_{j=1}^r \mathbf{e}_j^\top \left(\sum_{i=1}^n \mathbf{Z}(t_i, \boldsymbol{\theta}_0) - \mathbf{k} \right) \right| \tag{3.14}
\end{aligned}$$

where \mathbf{e}_j is the unit vector in the j th coordinate direction and

$$S = \frac{1}{n} \sum_{i=1}^{n-1} \mathbf{Z}(t_i, \boldsymbol{\theta}_0) \mathbf{Z}^\top(t_i, \boldsymbol{\theta}_0) = -\mathbf{h}'(\mathbf{0}, \boldsymbol{\theta}_0) \xrightarrow{P} \Sigma_Z.$$

Now $\boldsymbol{\phi}^\top S \boldsymbol{\phi} \geq \xi_r + o_p(1)$ where $\xi_r > 0$ is the smallest eigenvalue of Σ_Z . The second

term in (3.14) is $O_p(n^{-1/2})$ by Lemma 2 and the central limit theorem. It follows that

$$\frac{\rho}{1 + \rho B_n} = O_p(n^{-1/2}).$$

Notice that $1 + \boldsymbol{\lambda}(\boldsymbol{\theta})^\top \mathbf{Z}(t_i, \boldsymbol{\theta})$ is always positive, which is implied by (3.8), therefore by (3.13)

$$\rho = \|\boldsymbol{\lambda}(\boldsymbol{\theta}_0)\| = O_p(n^{-1/2}). \quad \blacksquare$$

3.5 Computational Considerations

Let us consider the simplest situation first, with only one parameter and two constraint equations.

Let X_1, X_2, \dots, X_n be i.i.d. non-negative random variables with hazard function Λ_0 and parameter θ . Let C_1, C_2, \dots, C_n be i.i.d. random censoring times. We only observe (t_i, δ_i) , where

$$t_i = \min(x_i, c_i) \quad \text{and} \quad \delta_i = I[x_i \leq c_i], \quad \text{for } i = 1, 2, \dots, n. \quad (3.15)$$

Let $w_i = \Delta\Lambda(t_i)$ for $i = 1, 2, \dots, n$. Again, without loss of generality, we assume the t_i s are already sorted in an increasing order, and we assume $w_n = \delta_n$. The log empirical likelihood we will study is

$$\log AL = \sum_{i=1}^n \delta_i \log w_i - \sum_{i=1}^n w_i R_i \quad (3.16)$$

where $R_i = \sum_j I[t_j \geq t_i]$.

Suppose information about θ is given in the following two equations:

$$\int g_1(t) d\Lambda_0(t) = \theta$$

$$\int g_2(t)d\Lambda_0(t) = \theta$$

where $g_1(t), g_2(t)$ are some functionally independent estimating functions that satisfy some moment conditions. Then we have the following two constraint equations:

$$\sum_{i=1}^{n-1} \delta_i g_1(t_i) w_i + g_1(t_n) w_n = \theta \quad (3.17)$$

$$\sum_{i=1}^{n-1} \delta_i g_2(t_i) w_i + g_2(t_n) w_n = \theta \quad (3.18)$$

Now the problem is how to maximize (3.16) under the above two constraints (3.17) and (3.18) and thereby determine an estimator $\hat{\theta}$ for θ .

It is well-known that the Nelson-Aalen estimator $\hat{w}_i = \delta_i/R_i$ achieves the global maximum for $\log AL$ in (3.16). If there is only one constraint equation (3.17), then the empirical-likelihood estimator $\tilde{\theta}$ achieves the constrained maximum for $\log AL$ in (3.16), where \tilde{w}_i maximizes (3.16) while at the same time satisfying (3.17).

If there are two constraint equations (3.17) and (3.18), however, it is more difficult to find a \hat{w}_i which maximizes (3.16) while at the same time satisfies both (3.17) and (3.18).

Pan and Zhou (2002) discussed the computational issue for the censored empirical likelihood with only one constraint equation. All they need to solve is the constraint equation for λ and it is monotone decreasing in λ . An Splus function that computes the empirical likelihood ratio described in that paper is available from the second author.

The computation of the over-determined maximum-empirical-likelihood estimator $\hat{\theta}$ is challenging. Let us consider this specific example as an illustration. In order to

evaluate $\log AL$ for a given θ , we must solve (3.10) for $w_i(\theta)$, which can often be accomplished by Newton's method. Next we propose to get the maximum-empirical-likelihood estimator $\tilde{\theta}$ for the just-determined case first, by solving the constraint equations (3.17) and (3.18) for θ , where $\hat{w}_i = \delta_i/R_i$. This will give us some information about where $\hat{\theta}$ could be. We reason that $\hat{\theta}$ is typically close to $\tilde{\theta}$, since both of them are \sqrt{n} consistent to the same true parameter θ_0 .

This idea works well in our simulation studies in Chapter 6. We obtain the just-determined maximum-empirical-likelihood-estimators $\tilde{\theta}_1$ and $\tilde{\theta}_2$ first, by solving the two constraint equations with $\hat{w}_i = \delta_i/R_i$, respectively. Then using Theorem 1, we calculate a series of $\max \log AL$ for different θ values that are close to $\tilde{\theta}_1$ and $\tilde{\theta}_2$. The θ value that corresponds to the maximization of these $\max \log AL$ is our over-determined maximum-empirical-likelihood estimator $\hat{\theta}$.

Computation becomes even more demanding when we have multiple parameters, say θ_1 and θ_2 , and multiple constraint equations. Any optimization technique can be tried, and further investigation is needed to see which works best.

Chapter 4 Hazard-Type Hypothesis for K-sample U-statistic

4.1 Introduction

A U-statistic is a class of non-parametric statistics that is especially important in estimation theory, which includes many important statistics such as sample mean and sample variance. The basic theory of U-statistic was developed by Wassily Hoeffding (1948) [11] and many others after that. In elementary statistics, U-statistics arise naturally in producing minimum-variance unbiased estimators. The theory related to U-statistics allows a single theoretical framework to be used to prove results relating to the asymptotic normality and to the variance (in finite samples) of a wide range of test-statistics and estimators. In addition the theory has applications to estimators which are not themselves U-statistics.

Let X_1, X_2, \dots, X_n be i.i.d. from an unknown population P in a nonparametric family \mathcal{P} . In a large class of problems, the parameter to be estimated is of the form

$$\theta = E[h(X_1, \dots, X_m)]$$

with a positive integer m and a Borel function h that is assumed to be symmetric and satisfies

$$E | h(X_1, \dots, X_m) | < \infty$$

for any $P \in \mathcal{P}$.

It is easy to see that a symmetric unbiased estimator of θ is

$$U_n = \binom{n}{m}^{-1} \sum_c h(X_{i_1}, \dots, X_{i_m}),$$

where \sum_c denotes the summation over the $\binom{n}{m}$ combinations of m distinct elements $\{i_1, \dots, i_m\}$ from $\{1, \dots, n\}$. The statistic U_n above is called a U-statistic with kernel h of order m .

Owen (2001) [26] investigated the two-sample U-statistic of order $(1, 1)$ with empirical likelihood. Owen's result made it possible to use empirical likelihood to construct confidence intervals based on a two-sample U-statistic, rather than using asymptotic normality to construct the confidence intervals.

Let X_1, X_2, \dots, X_m be i.i.d. from an unknown population P_1 and Y_1, Y_2, \dots, Y_n be i.i.d. from an unknown population P_2 . The parameter to be estimated is of the form

$$\theta = E[h(X_1, Y_1)].$$

The two-sample U-statistic Owen studied is

$$U_{mn} = \sum_{i=1}^m \sum_{j=1}^n h(x_i, y_j) \frac{1}{m} \frac{1}{n}, \quad (4.1)$$

which is an unbiased estimator of θ .

In this chapter we will extend Owen's result and study the use of empirical likelihood with a k -sample U-statistic of order $(1, 1, \dots, 1)$ in terms of hazard function, where k is a fixed positive integer and greater than 1. We will give a k -sample theorem below. Moreover, we will also allow right censoring in the data. The details of the proof are given only for the two-sample case but the k -sample case can be similarly proved. The proof for the two-sample case is collaborative effort of Dr. Bill Barton and myself, under the supervision of our advisor Dr. Mai Zhou. Integral signs without explicit limits are understood to encompass the entire support of the

variable, here and elsewhere.

Suppose that $x_{11}, x_{12}, \dots, x_{1n_1}$ are i.i.d. non-negative observations of lifetimes with a continuous distribution function F_1 and cumulative hazard function Λ_1 . Independent of the lifetimes there are censoring times $c_{11}, c_{12}, \dots, c_{1n_1}$ which are i.i.d. with a distribution G_1 . In practice F_1 , Λ_1 and G_1 will be unknown. Only the censored observations, (t_{1i}, d_{1i}) , are available to us:

$$t_{1i} = \min(x_{1i}, c_{1i}) \quad \text{and} \quad d_{1i} = I[x_{1i} \leq c_{1i}] \quad \text{for } i = 1, 2, \dots, n_1 \quad (4.2)$$

Similarly defined with $F_2, \Lambda_2, G_2, \dots, F_k, \Lambda_k, G_k$, we have

$$t_{2i} = \min(x_{2i}, c_{2i}) \quad \text{and} \quad d_{2i} = I[x_{2i} \leq c_{2i}] \quad \text{for } i = 1, 2, \dots, n_2 \quad (4.3)$$

.....

$$t_{ki} = \min(x_{ki}, c_{ki}) \quad \text{and} \quad d_{ki} = I[x_{ki} \leq c_{ki}] \quad \text{for } i = 1, 2, \dots, n_k \quad (4.4)$$

Definition: The U-statistic with respect to hazard is defined as

$$\hat{\theta} = \sum_{i_1=1}^{n_1} \dots \sum_{i_k=1}^{n_k} H(t_{1i_1}, \dots, t_{ki_k}) \hat{w}_{1i_1} \dots \hat{w}_{ki_k}$$

where $\hat{w}_{1i_1} \dots \hat{w}_{ki_k}$ are the jumps of the Nelson-Aalen estimators based on $(t_{11}, d_{11}), \dots, (t_{1n_1}, d_{1n_1}); (t_{21}, d_{21}), \dots, (t_{2n_2}, d_{2n_2}); \dots; (t_{k1}, d_{k1}), \dots, (t_{kn_k}, d_{kn_k})$. This U-statistic is defined similarly to (4.1) and is an unbiased estimator of

$$\theta = \int \dots \int H(t_1, t_2, \dots, t_k) d\Lambda_1(t_1) d\Lambda_2(t_2) \dots d\Lambda_k(t_k).$$

Theorem: Suppose $(t_{11}, d_{11}), \dots, (t_{1n_1}, d_{1n_1}); (t_{21}, d_{21}), \dots, (t_{2n_2}, d_{2n_2}); \dots; (t_{k1}, d_{k1}), \dots, (t_{kn_k}, d_{kn_k})$ are from the above data setting. Suppose further that there is a hypothesized constraint of the following form:

$$H_o : \int \dots \int H(t_1, t_2, \dots, t_k) d\Lambda_1(t_1) d\Lambda_2(t_2) \dots d\Lambda_k(t_k) = \theta \quad (4.5)$$

where $H(t_1, t_2, \dots, t_k)$ is a continuous function and assumed to be bounded.

If $\int \dots \int H(t_1, t_2, \dots, t_k) d\Lambda_1(t_1) d\Lambda_2(t_2) \dots d\Lambda_k(t_k) < \infty$, and some further regularity conditions are satisfied, then under H_o , as $\min(n_1, n_2, \dots, n_k) \rightarrow \infty$, the distribution of the empirical log-likelihood ratio (ELLR) has the limit

$$-2ELLR \xrightarrow{\mathcal{D}} \chi_{(1)}^2 ,$$

where

$$ELLR = \max_{\Lambda_1, \dots, \Lambda_k, H_0} \log EL_k(\Lambda_1, \dots, \Lambda_k) - \log EL_k(\hat{\Lambda}_1, \dots, \hat{\Lambda}_k),$$

and $\log EL_k$ is defined in (4.14); and $\hat{\Lambda}_i, i = 1, \dots, k$ are the Nelson-Aalen estimators of $\Lambda_i, i = 1, \dots, k$.

We will give the theorem and proof below for the $k = 2$ case .

Theorem(Hu and Barton): Suppose $(s_1, dx_1), \dots, (s_n, dx_n), (t_1, dy_1), \dots, (t_m, dy_m)$ are observations as in (4.2) and (4.3). Suppose further that there is a hypothesized constraint of the following form:

$$H_o : \iint H(s, t) d\Lambda_X(s) d\Lambda_Y(t) = \theta \tag{4.6}$$

where $\Lambda_X(s)$ and $\Lambda_Y(t)$ are the respective cumulative-hazard functions of X and Y ; and $H(s, t)$ is a continuous function and assumed to be bounded.

If $\iint H(s, t) d\Lambda_X(s) d\Lambda_Y(t) < \infty$ and some further regularity conditions are satisfied as in the proof later, then under H_o , as $\min(n, m) \rightarrow \infty$, the distribution of the empirical log-likelihood ratio (ELLR) defined in section 4.3 has the limit

$$-2ELLR \xrightarrow{\mathcal{D}} \chi_{(1)}^2 .$$

Hazard functions are typically monotone-increasing and unbounded. Therefore the function $H(s, t)$ in (4.6) must approach 0 as $s, t \rightarrow \infty$. Two examples of such an

$H(s, t)$ are $e^{-|s+t|}$, and $e^{-|st|}$.

In order to evaluate H_o in the two-sample theorem we will calculate a constraint, similar in form to (4.6), but based on the data. The expression for this constraint is as follows:

$$\sum_{i=1}^n \sum_{j=1}^m H_{ij} \widehat{w}_i \widehat{v}_j - \theta = 0 \quad (4.7)$$

where \widehat{w}_i and \widehat{v}_j are the estimated jumps in hazard at x_i and y_j , respectively; and H_{ij} is an abbreviated notation for $H(x_i, y_j)$.

If we take the integrals in (4.6) to be Stieltjes integrals then (4.6) and (4.7) are fundamentally similar, only (4.6) applies to continuous (s, t) whereas (4.7) applies to discrete (s, t) .

We organize the proof of the two-sample Theorem into five sections. In section 4.2 we calculate maximum-likelihood estimates (MLE's) of the hazard jumps for each of the two samples. In section 4.3 we calculate the empirical-log-likelihood-ratio (ELLR) for the two samples. In section 4.4 we show that $-2ELLR$ is asymptotically distributed as $\chi_{(1)}^2$ under H_o . In section 4.5 we demonstrate that the Taylor expansions used in the calculations are valid. In section 4.6 we prove two lemmas used in the previous sections.

4.2 MLE's for the Hazard Jumps

In this section we calculate the maximum-likelihood-estimators (MLE's) for the hazard jumps in (4.6) and (4.7).

The calculation of the MLE's for the hazard jumps generally follows the derivation in Owen (2001) [26] pp. 223-227. There are some differences, however, since Owen's derivation involves a mean-type hypothesis for uncensored data, whereas our derivation involves a hazard-type hypothesis for right-censored data. Since the calculations are based on likelihood, we first discuss the likelihood expression that we will employ.

For the general k -sample case, the empirical likelihoods EL_{X_1} , EL_{X_2} , ..., EL_{X_k} for samples \mathbf{x}_1 , \mathbf{x}_2 , ..., \mathbf{x}_k can be written as follows:

$$EL_{X_1} = \prod_{i=1}^{n_1} w_{1i}^{d_{1i}} \exp \left(- \sum_{r=1}^{n_1} w_{1r} I[x_{1r} \leq x_{1i}] \right) \quad (4.8)$$

$$EL_{X_2} = \prod_{i=1}^{n_2} w_{2i}^{d_{2i}} \exp \left(- \sum_{r=1}^{n_2} w_{2r} I[x_{2r} \leq x_{2i}] \right) \quad (4.9)$$

.....

$$EL_{X_k} = \prod_{i=1}^{n_k} w_{ki}^{d_{ki}} \exp \left(- \sum_{r=1}^{n_k} w_{kr} I[x_{kr} \leq x_{ki}] \right) \quad (4.10)$$

The likelihood expression as in (4.8), (4.9), ..., (4.10) is described in Murphy (1995) [20]. It is a "Poisson extension" of the usual likelihood. Although it is not a true likelihood, it does yield a meaningful likelihood-ratio test and it has the advantage that the log-likelihood is easily formed from it. See also Pan and Zhou(2002) [29] for a brief discussion of this Poisson extension.

The log-empirical-likelihoods ($\log EL_{X_1}$, $\log EL_{X_2}$, ..., $\log EL_{X_k}$ for the k samples can then be expressed as

$$\log EL_{X_1} = \sum_{i=1}^{n_1} \left(d_{1i} \log w_{1i} - \sum_{r=1}^{n_1} w_{1r} I[x_{1r} \leq x_{1i}] \right) \quad \text{from (4.8)} \quad (4.11)$$

$$\log EL_{X_2} = \sum_{i=1}^{n_2} \left(d_{2i} \log w_{2i} - \sum_{r=1}^{n_2} w_{2r} I[x_{2r} \leq x_{2i}] \right) \quad \text{from (4.9)} \quad (4.12)$$

.....

$$\log EL_{X_k} = \sum_{i=1}^{n_k} \left(d_{ki} \log w_{ki} - \sum_{r=1}^{n_k} w_{kr} I[x_{kr} \leq x_{ki}] \right) \quad \text{from (4.10)} \quad (4.13)$$

Since the k samples are independent, therefore the log-empirical-likelihood for the set of all k samples, denoted as $\log EL_k$, can be expressed as the sum of (4.11), (4.12), and (4.13) as follows:

$$\log EL_k = \log EL_{X_1} + \log EL_{X_2} + \dots + \log EL_{X_k}. \quad (4.14)$$

We will give the detailed proof for the two-sample case below. The empirical likelihoods EL_X and EL_Y for the two samples \mathbf{x} and \mathbf{y} can be written as follows:

$$EL_X = \prod_{i=1}^n w_i^{dx_i} \exp \left(- \sum_{r=1}^n w_r I[x_r \leq x_i] \right) \quad (4.15)$$

$$EL_Y = \prod_{j=1}^m \nu_j^{dy_j} \exp \left(- \sum_{s=1}^m \nu_s I[y_s \leq y_j] \right) \quad (4.16)$$

The log-empirical-likelihoods ($\log EL_X$ and $\log EL_Y$) for the two samples can then be expressed as

$$\log EL_X = \sum_{i=1}^n \left(dx_i \log w_i - \sum_{r=1}^n w_r I[x_r \leq x_i] \right) \quad \text{from (4.15)} \quad (4.17)$$

$$\log EL_Y = \sum_{j=1}^m \left(dy_j \log \nu_j - \sum_{s=1}^m \nu_s I[y_s \leq y_j] \right) \quad \text{from (4.16)}. \quad (4.18)$$

The log-empirical-likelihood for the set of both samples, denoted as $\log EL$, can be expressed as the sum of (4.17) and (4.18) as follows:

$$\log EL = \log EL_X + \log EL_Y. \quad (4.19)$$

The maximization of the unconstrained likelihood in (4.17) is easily accomplished by taking a partial derivative with respect to w_i and then equating the partial derivative to zero. The maximization of the unconstrained likelihood in (4.18) is accomplished similarly. Expressions for the respective hazard-jump MLE's \tilde{w} and $\tilde{\nu}_j$ are as follows:

$$\tilde{w}_i = \frac{dx_i}{Rx_i}, i = 1, \dots, n \quad (4.20)$$

$$\tilde{\nu}_j = \frac{dy_j}{Ry_j}, j = 1, \dots, m. \quad (4.21)$$

where $\mathbf{dx} = (dx_1, \dots, dx_n)$ and $\mathbf{dy} = (dy_1, \dots, dy_m)$ are the respective censoring indicators for \mathbf{x} and \mathbf{y} . For \mathbf{x} we use the convention that $dx_i = 1$ for an uncensored datum and $dx_i = 0$ for a right-censored datum, and similarly for \mathbf{y} . Rx_i and Ry_j are the number of survivors at x_i^- and y_j^- respectively, defined as,

$$Rx_i = \sum_{r=1}^n I[x_r \geq x_i] \quad \text{and} \quad Ry_j = \sum_{s=1}^m I[y_s \geq y_j]. \quad (4.22)$$

The estimators in (4.20) and (4.21) are commonly referred to as Nelson-Aalen estimators (e.g. Kalbfleisch and Prentice [13] p. 18). They are sometimes notated as $d\widehat{\Lambda}_X(x_i)$ and $d\widehat{\Lambda}_Y(y_j)$, respectively, and we will adopt this latter notation in section 4.4.

The calculation of the MLE's for the constrained hazard jumps, which is considerably more complicated, will comprise the rest of this section 4.2.

We use $\log EL$ in (4.19) and the constraint in (4.7) to construct a constrained-log-likelihood target-function G . The constraint in (4.7) is incorporated into G by means of a Lagrange multiplier λ . The result is as follows:

$$G = \log EL - \lambda \left(\sum_{i=1}^n \sum_{j=1}^m H_{ij} w_i \nu_j - \theta \right). \quad (4.23)$$

We calculate the constrained MLE's of w_i , ν_j , and λ by taking partial derivatives of G with respect to w_i , ν_j , and λ and then equating these partial derivatives to zero. We denote the resulting three constrained MLE's as \widehat{w}_i , $\widehat{\nu}_j$, and $\widehat{\lambda}$, respectively.

In the course of calculating the constrained MLE's of w_i , ν_j and λ we will make use of the following two quantities:

$$\widetilde{H}_{i.} = \sum_{j=1}^m \widehat{\nu}_j H_{ij} \quad \text{and} \quad \widetilde{H}_{.j} = \sum_{i=1}^n \widehat{w}_i H_{ij}. \quad (4.24)$$

Note that $\widetilde{H}_{i.}$ and $\widetilde{H}_{.j}$ are uniformly bounded since H_{ij} is necessarily uniformly

bounded per (4.6) above.

We proceed as follows:

$$\frac{\partial G}{\partial w_i} \Big|_{w_i=\hat{w}_i, \nu_j=\hat{\nu}_j, \lambda=\hat{\lambda}} = 0. \quad (4.25)$$

Combining (4.25), (4.17)-(4.19) and (4.23) gives

$$\frac{dx_i}{\hat{w}_i} - \sum_{r=1}^n I[x_r \geq x_i] - \hat{\lambda} \sum_{j=1}^m H_{ij} \hat{\nu}_j = 0. \quad (4.26)$$

Rearranging (4.26) gives

$$\hat{w}_i = \frac{dx_i}{\left(Rx_i + \hat{\lambda} \tilde{H}_i\right)} \quad (4.27)$$

using (4.22) and (4.24).

Similarly we set

$$\frac{\partial G}{\partial \nu_j} \Big|_{w_i=\hat{w}_i, \nu_j=\hat{\nu}_j, \lambda=\hat{\lambda}} = 0. \quad (4.28)$$

Combining (4.28), (4.17)-(4.19) and (4.23) gives

$$\frac{dy_j}{\hat{\nu}_j} - \sum_{s=1}^m I[y_s \geq y_j] - \hat{\lambda} \sum_{i=1}^n H_{ij} \hat{w}_i = 0 \quad (4.29)$$

Rearranging (4.29) gives

$$\hat{\nu}_j = \frac{dy_j}{\left(Ry_j + \hat{\lambda} \tilde{H}_{.j}\right)}. \quad (4.30)$$

using (4.22) and (4.24).

The second derivatives $\frac{\partial^2 G}{\partial w_i^2}$ and $\frac{\partial^2 G}{\partial v_j^2}$ are easily shown to be negative, confirming that \widehat{w}_i and \widehat{v}_j are indeed maxima.

Lastly we set

$$\left. \frac{\partial G}{\partial \lambda} \right|_{w_i=\widehat{w}_i, v_j=\widehat{v}_j, \lambda=\widehat{\lambda}} = 0. \quad (4.31)$$

Combining (4.31), (4.17)-(4.19) and (4.23) gives

$$\sum_{i=1}^n \sum_{j=1}^m H_{ij} \widehat{w}_i \widehat{v}_j = \theta \quad (4.32)$$

which is just the constraint as in (4.7).

Equations (4.27), (4.30), and (4.32) must be solved simultaneously to find explicit expressions for \widehat{w}_i , \widehat{v}_j , and $\widehat{\lambda}$. A precise solution would require numerical methods. We instead calculate an approximate solution using Taylor expansions. We begin by finding an approximate value of $\widehat{\lambda}$.

First we expand \widehat{w}_i and \widehat{v}_j in (4.27) and (4.30) in order to bring $\widehat{\lambda}$ into the numerator. This results in the following two expressions:

$$\widehat{w}_i = \frac{dx_i}{Rx_i} \left[1 - \left(\frac{\widehat{\lambda}}{Rx_i} \widetilde{H}_i \right) + \left(\frac{\widehat{\lambda}}{Rx_i} \widetilde{H}_i \right)^2 - \dots \right] \quad (4.33)$$

$$\widehat{v}_j = \frac{dy_j}{Ry_j} \left[1 - \left(\frac{\widehat{\lambda}}{Ry_j} \widetilde{H}_{.j} \right) + \left(\frac{\widehat{\lambda}}{Ry_j} \widetilde{H}_{.j} \right)^2 - \dots \right] \quad (4.34)$$

where we have used the Taylor expansion

$$\frac{1}{1+\epsilon} = 1 - \epsilon + \epsilon^2 - \dots, \quad \text{valid for } |\epsilon| < 1. \quad (4.35)$$

Later in section 4.5 we will show that

$$\frac{\hat{\lambda}}{Rx_i} \tilde{H}_i \xrightarrow{p} 0 \quad \text{as } \min(n, m) \rightarrow \infty \quad (4.36)$$

$$\frac{\hat{\lambda}}{Ry_j} \tilde{H}_j \xrightarrow{p} 0 \quad \text{as } \min(n, m) \rightarrow \infty \quad (4.37)$$

so that the restriction in (4.35) will hold when $\min(n, m)$ is sufficiently large. Also, (4.36) and (4.37) will allow us to drop the higher-order terms in (4.33) and (4.34) as asymptotically negligible.

To simplify the notation in (4.33) and (4.34) let

$$\eta_i = \frac{\hat{\lambda}}{Rx_i} \tilde{H}_i. \quad (4.38)$$

$$\kappa_j = \frac{\hat{\lambda}}{Ry_j} \tilde{H}_j. \quad (4.39)$$

Then (4.33) and (4.34) can be written as

$$\hat{w}_i = \frac{dx_i}{Rx_i} \left(1 - \eta_i + \eta_i^2 - \dots \right) \quad (4.40)$$

$$\hat{v}_j = \frac{dy_j}{Ry_j} \left(1 - \kappa_j + \kappa_j^2 - \dots \right) \quad (4.41)$$

where

$$\eta_i \xrightarrow{p} 0 \quad \text{as } \min(n, m) \rightarrow \infty, \quad \text{from (4.36) and (4.38)} \quad (4.42)$$

$$\kappa_j \xrightarrow{p} 0 \quad \text{as } \min(n, m) \rightarrow \infty, \quad \text{from (4.37) and (4.39)}. \quad (4.43)$$

Substituting (4.40) and (4.41) into (4.32) gives the following:

$$\theta = \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{Rx_i} \frac{dy_j}{Ry_j} \left(1 - \eta_i + \eta_i^2 - \dots\right) \left(1 - \kappa_j + \kappa_j^2 - \dots\right) \quad (4.44)$$

$$\begin{aligned} &= \bar{H}_{..} - \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{Rx_i} \frac{dy_j}{Ry_j} (\eta_i + \kappa_j) \\ &\quad + \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{Rx_i} \frac{dy_j}{Ry_j} (\eta_i^2 + \kappa_j^2 + \eta_i \kappa_j) - \dots \end{aligned} \quad (4.45)$$

$$\text{where } \bar{H}_{..} = \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{Rx_i} \frac{dy_j}{Ry_j} \quad (4.46)$$

$$\doteq \bar{H}_{..} - \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{Rx_i} \frac{dy_j}{Ry_j} (\eta_i + \kappa_j), \quad (4.47)$$

dropping higher order terms per (4.42) and (4.43).

Now η_i and κ_j in the right-hand side (RHS) of (4.47) are themselves functions of $\hat{\lambda}$, per (4.38) and (4.39) above. In order to facilitate the objective of finding an explicit approximation for $\hat{\lambda}$ we introduce the terms $\bar{H}_{.i}$ and $\bar{H}_{.j}$, which are not functions of $\hat{\lambda}$, as follows:

$$\bar{H}_{.i} = \sum_{j=1}^m H_{ij} \frac{dy_j}{Ry_j} \quad \text{and} \quad \bar{H}_{.j} = \sum_{i=1}^n H_{ij} \frac{dx_i}{Rx_i}. \quad (4.48)$$

We obtain (4.48) by substituting $\frac{dy_j}{Ry_j}$ for $\hat{\nu}_j$ and $\frac{dx_i}{Rx_i}$ for $\hat{\mu}_i$ in (4.24). Note that $\bar{H}_{.i}$

and $\bar{H}_{.j}$ are uniformly bounded since H_{ij} is assumed to be uniformly bounded per (4.6) above.

Now consider the following:

$$\tilde{H}_i = \bar{H}_i + (\tilde{H}_i - \bar{H}_i) \quad (4.49)$$

$$= \bar{H}_i + \sum_{j=1}^m H_{ij} \left(\hat{\nu}_j - \frac{dy_j}{Ry_j} \right) \quad \text{from (4.24) and (4.48)} \quad (4.50)$$

$$= \bar{H}_i + \sum_{j=1}^m H_{ij} \left[\frac{dy_j}{Ry_j} (1 - \kappa_j + \kappa_j^2 - \dots) - \frac{dy_j}{Ry_j} \right] \quad \text{from (4.41)} \quad (4.51)$$

$$= \bar{H}_i - \sum_{j=1}^m H_{ij} \frac{dy_j}{Ry_j} (\kappa_j - \kappa_j^2 + \dots) \quad (4.52)$$

$$= \bar{H}_i - \sum_{j=1}^m H_{ij} \frac{dy_j}{Ry_j} o_p(1) \quad \text{from (4.43), assuming } \max_{1 \leq j \leq m} (\kappa_j) = o_p(1) \quad (4.53)$$

$$= (1 - o_p(1)) \bar{H}_i \quad \text{from (4.48)}. \quad (4.54)$$

Similarly we can calculate

$$\tilde{H}_{.j} = (1 - o_p(1)) \bar{H}_{.j} \quad \text{assuming } \max_{1 \leq i \leq n} (\eta_i) = o_p(1). \quad (4.55)$$

Then we can estimate $\hat{\lambda}$ as follows:

$$\theta \doteq \bar{H}_{..} - \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{Rx_i} \frac{dy_j}{Ry_j} (\eta_i + \kappa_j) \quad \text{from (4.47)} \quad (4.56)$$

$$= \bar{H}_{..} - \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{Rx_i} \frac{dy_j}{Ry_j} \left(\frac{\hat{\lambda}}{Rx_i} \tilde{H}_i + \frac{\hat{\lambda}}{Ry_j} \tilde{H}_{.j} \right) \quad \text{from (4.38), (4.39)} \quad (4.57)$$

$$= \bar{H}_{..} - \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{Rx_i} \frac{dy_j}{Ry_j} \left[\frac{\hat{\lambda}}{Rx_i} \bar{H}_{i.} (1 - o_p(1)) + \frac{\hat{\lambda}}{Ry_j} \bar{H}_{.j} (1 - o_p(1)) \right] \quad (4.58)$$

from (4.54), (4.55)

$$= \bar{H}_{..} - \hat{\lambda} (1 - o_p(1)) \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{Rx_i} \frac{dy_j}{Ry_j} \left(\frac{\bar{H}_{i.}}{Rx_i} + \frac{\bar{H}_{.j}}{Ry_j} \right) \quad (4.59)$$

$$\doteq \bar{H}_{..} - \hat{\lambda} \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{Rx_i} \frac{dy_j}{Ry_j} \left(\frac{\bar{H}_{i.}}{Rx_i} + \frac{\bar{H}_{.j}}{Ry_j} \right). \quad (4.60)$$

Then from (4.60) the desired approximation of λ is,

$$\hat{\lambda} \doteq (\bar{H}_{..} - \theta) / D \quad (4.61)$$

$$\text{where } D = \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{Rx_i} \frac{dy_j}{Ry_j} \left(\frac{\bar{H}_{i.}}{Rx_i} + \frac{\bar{H}_{.j}}{Ry_j} \right) \quad (4.62)$$

The RHS of (4.61) does not involve \hat{w}_i, \hat{v}_j , or $\hat{\lambda}$, rather it only involves the data. Therefore the RHS of (4.61) is an explicit approximation of $\hat{\lambda}$.

By a similar argument we can substitute (4.54) and (4.55) into (4.33) and (4.34), respectively. This gives an approximation \check{w}_i for \hat{w}_i and an approximation \check{v}_j for \hat{v}_j , as follows:

$$\check{w}_i = \frac{dx_i}{Rx_i} \left(1 - \frac{\hat{\lambda}}{Rx_i} \bar{H}_{i.} \right) \doteq \hat{w}_i \quad (4.63)$$

$$\check{v}_j = \frac{dy_j}{Ry_j} \left(1 - \frac{\hat{\lambda}}{Ry_j} \bar{H}_{.j} \right) \doteq \hat{v}_j. \quad (4.64)$$

4.3 ELLR for the Two Samples

In this section we calculate an expression for the empirical-log-likelihood-ratio (ELLR) for the two samples under the constraint in (4.6).

Let us define the following:

$$\check{\eta}_i = \frac{\hat{\lambda}}{Rx_i} \bar{H}_i, \quad i = 1, \dots, n \quad \text{similar to (4.38)} \quad (4.65)$$

$$\check{\boldsymbol{\eta}} = (\check{\eta}_1, \check{\eta}_2, \dots, \check{\eta}_n) \quad (4.66)$$

$$Q(\check{\boldsymbol{\eta}}) = \sum_{i=1}^n \left(dx_i \log \left(\frac{dx_i}{Rx_i} (1 - \check{\eta}_i) \right) - \sum_{r=1}^n \frac{dx_r}{Rx_r} (1 - \check{\eta}_r) I[x_r \leq x_i] \right) \quad (4.67)$$

$$Q_o = Q(\mathbf{0}) = \sum_{i=1}^n \left(dx_i \log \left(\frac{dx_i}{Rx_i} \right) - \sum_{r=1}^n \frac{dx_r}{Rx_r} I[x_r \leq x_i] \right). \quad (4.68)$$

From (4.17), (4.63), and (4.65) we see that $Q(\check{\boldsymbol{\eta}})$ in (4.67) is an estimator for the constrained log-likelihood for the sample \mathbf{x} and we see that Q_o in (4.68) is an estimator for the unconstrained log-likelihood for the sample \mathbf{x} .

Similarly let us define the following:

$$\check{\kappa}_j = \frac{\hat{\lambda}}{Ry_j} \bar{H}_j, \quad j = 1, \dots, m \quad \text{similar to (4.39)} \quad (4.69)$$

$$\check{\boldsymbol{\kappa}} = (\check{\kappa}_1, \check{\kappa}_2, \dots, \check{\kappa}_m) \quad (4.70)$$

$$P(\check{\boldsymbol{\kappa}}) = \sum_{j=1}^m \left(dy_j \log \left(\frac{dy_j}{Ry_j} (1 - \check{\kappa}_j) \right) - \sum_{s=1}^m \frac{dy_s}{Ry_s} (1 - \check{\kappa}_s) I[y_s \leq y_j] \right) \quad (4.71)$$

$$P_o = P(\mathbf{0}) = \sum_{j=1}^m \left(dy_j \log \left(\frac{dy_j}{Ry_j} \right) - \sum_{s=1}^m \frac{dy_s}{Ry_s} I[y_s \leq y_j] \right). \quad (4.72)$$

From (4.18), (4.64), and (4.69) we see that $P(\check{\boldsymbol{\kappa}})$ is an estimator for the constrained log-likelihood for the sample \mathbf{y} and we see that P_o is an estimator for the uncon-

strained log-likelihood for the sample \mathbf{y} .

It is convenient to rewrite $Q(\check{\eta})$ in (4.67) as follows:

$$Q(\check{\eta}) = \sum_{i=1}^n \left(dx_i \log \left(\frac{dx_i}{Rx_i} (1 - \check{\eta}_i) \right) \right) - \sum_{i=1}^n \sum_{r=1}^n \left(\frac{dx_r}{Rx_r} (1 - \check{\eta}_r) I[x_r \leq x_i] \right) \quad (4.73)$$

$$= \sum_{i=1}^n \left(dx_i \log \left(\frac{dx_i}{Rx_i} (1 - \check{\eta}_i) \right) \right) - \sum_{r=1}^n \left(\frac{dx_r}{Rx_r} (1 - \check{\eta}_r) \sum_{i=1}^n I[x_r \leq x_i] \right) \quad (4.74)$$

$$= \sum_{i=1}^n \left(dx_i \log \left(\frac{dx_i}{Rx_i} (1 - \check{\eta}_i) \right) \right) - \sum_{r=1}^n \left(dx_r (1 - \check{\eta}_r) \right) \quad \text{using (4.22)} \quad (4.75)$$

$$= \sum_{i=1}^n \left(dx_i \log \left(\frac{dx_i}{Rx_i} (1 - \check{\eta}_i) \right) - dx_i (1 - \check{\eta}_i) \right) \quad \text{changing index r to i} \quad (4.76)$$

$$= \sum_{i=1}^n Q_i(\check{\eta}_i) \quad (4.77)$$

$$\text{where } Q_i(\check{\eta}_i) = \left(dx_i \log \left(\frac{dx_i}{Rx_i} (1 - \check{\eta}_i) \right) - dx_i (1 - \check{\eta}_i) \right). \quad (4.78)$$

By a similar calculation we can rewrite $P(\check{\kappa})$ in (4.71) as follows:

$$P(\check{\kappa}) = \sum_{j=1}^m P_j(\check{\kappa}_j) \quad (4.79)$$

$$\text{where } P_i(\check{\kappa}_j) = \left(dy_j \log \left(\frac{dy_j}{Ry_j} (1 - \check{\kappa}_j) \right) - dy_j (1 - \check{\kappa}_j) \right). \quad (4.80)$$

We will require the quantities $\sum_{i=1}^n \check{\eta}_i Q'_i(0)$, $\sum_{j=1}^m \check{\kappa}_j P'_j(0)$, $\sum_{i=1}^n \frac{\check{\eta}_i^2}{2} Q''_i(0)$, and $\sum_{j=1}^m \frac{\check{\kappa}_j^2}{2} P''_j(0)$ which we calculate as follows:

$$\sum_{i=1}^n \check{\eta}_i Q'_i(0) = \sum_{i=1}^n \check{\eta}_i \left(\frac{-dx_i \frac{dx_i}{R x_i}}{\frac{dx_i}{R x_i} (1 - \check{\eta}_i)} + dx_i \right) \Big|_{\check{\eta}_i=0} \quad \text{using (4.78)} \quad (4.81)$$

$$= 0 \quad (4.82)$$

A similar calculation shows

$$\sum_{j=1}^m \check{\kappa}_j P'_j(0) = 0 \quad \text{using (4.80)} \quad (4.83)$$

And,

$$\sum_{i=1}^n \frac{\check{\eta}_i^2}{2} Q''_i(0) = \sum_{i=1}^n \frac{\check{\eta}_i^2}{2} \left(\frac{-dx_i}{(1 - \eta_i)^2} \right) \Big|_{\check{\eta}_i=0} \quad \text{using (4.81)} \quad (4.84)$$

$$= - \sum_{i=1}^n \frac{\check{\eta}_i^2}{2} dx_i \quad (4.85)$$

A similar calculation shows

$$\sum_{j=1}^m \frac{\check{\kappa}_j^2}{2} P''_j(0) = - \sum_{j=1}^m \frac{\check{\kappa}_j^2}{2} dy_j \quad (4.86)$$

Then we can calculate $-2ELLR$ as follows:

$$-2ELLR \doteq -2(Q(\check{\eta}) - Q_o) - 2(P(\check{\kappa}) - P_o) \quad (4.87)$$

from (4.17) – (4.19), (4.65) – (4.72)

$$= -2Q(\check{\eta}) + 2Q_o - 2P(\check{\kappa}) + 2P_o \quad (4.88)$$

$$\begin{aligned} = & -2\left(Q_o + \sum_{i=1}^n \left(\check{\eta}_i Q'_i(0) + \frac{\check{\eta}_i^2}{2} Q''_i(0)\right)\right) + 2Q_o + R_n(\check{\eta}) \\ & - 2\left(P_o + \sum_{j=1}^m \left(\check{\kappa}_j P'_j(0) + \frac{\check{\kappa}_j^2}{2} P''_j(0)\right)\right) + 2P_o + R_m(\check{\kappa}) \end{aligned} \quad (4.89)$$

by Taylor expansion of $Q(\check{\eta})$ and $P(\check{\kappa})$ about $\mathbf{0}$,

where $R_n(\check{\eta})$ and $R_m(\check{\kappa})$ are the remainder terms

$$\doteq -\sum_{i=1}^n \check{\eta}_i^2 Q''_i(0) - \sum_{j=1}^m \check{\kappa}_j^2 P''_j(0) \quad (4.90)$$

using (4.82), (4.83), dropping remainder terms

$$= \sum_{i=1}^n \check{\eta}_i^2 dx_i + \sum_{j=1}^m \check{\kappa}_j^2 dy_j \quad \text{using (4.85), (4.86)} \quad (4.91)$$

$$= \hat{\lambda}^2 \sum_{i=1}^n \left(\frac{\overline{H}_i}{R x_i}\right)^2 dx_i + \hat{\lambda}^2 \sum_{j=1}^m \left(\frac{\overline{H}_j}{R y_j}\right)^2 dy_j \quad \text{using (4.65), (4.69)} \quad (4.92)$$

$$= \hat{\lambda}^2 \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dy_j}{R y_j} \frac{dx_i}{R x_i} \frac{\overline{H}_i}{R x_i} + \hat{\lambda}^2 \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{R x_i} \frac{dy_j}{R y_j} \frac{\overline{H}_j}{R y_j} \quad (4.93)$$

using (4.48)

$$= \widehat{\lambda}^2 D \text{ from (4.62)} \quad (4.94)$$

$$\doteq \frac{(\overline{H}_{..} - \theta)^2}{D} \text{ from (4.61)} \quad (4.95)$$

4.4 Distribution of -2ELLR

In this section we show that the distribution of -2ELLR goes asymptotically to χ_1^2 as $\min(n, m) \rightarrow \infty$, when the null hypothesis H_o is true.

Let us establish the following notation:

$$\text{Summations are represented as Stieltjes integrals} \quad (4.96)$$

$$x_i \text{ is represented as } t \quad (4.97)$$

$$y_j \text{ is represented as } s \quad (4.98)$$

$$H_{ij} \text{ is represented as } g(t, s) \quad (4.99)$$

$$\frac{dx_i}{Rx_i} \text{ is represented as } d\widehat{\Lambda}_1(t), \text{ a Nelson-Aalen jump as in (4.20)} \quad (4.100)$$

$$\frac{dy_j}{Ry_j} \text{ is represented as } d\widehat{\Lambda}_2(s), \text{ a Nelson-Aalen jump as in (4.21)} \quad (4.101)$$

$$H_o \text{ in (4.6) can then be denoted as } \iint g(t, s) d\Lambda_1(t) d\Lambda_2(s) = \theta. \quad (4.102)$$

The derivation below uses Martingale theory and the reader will need to be familiar with this theory in order to follow the derivation. A good reference for Martingale theory is Kalbfleisch and Prentice (2002) [13], chapter 5.

Now consider,

$$\sqrt{\frac{nm}{n+m}}(\bar{H}_{..} - \theta) = \sqrt{\frac{nm}{n+m}} \left(\sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{Rx_i} \frac{dy_j}{Ry_j} - \theta \right) \quad (4.103)$$

from (4.46)

$$= \sqrt{\frac{nm}{n+m}} \left(\iint g(t, s) d\hat{\Lambda}_1(t) d\hat{\Lambda}_2(s) - \iint g(t, s) d\Lambda_1(t) d\Lambda_2(s) \right) \quad (4.104)$$

using (4.96) to (4.102)

$$\begin{aligned} &= \sqrt{\frac{nm}{n+m}} \left(\iint g(t, s) d\hat{\Lambda}_1(t) d\hat{\Lambda}_2(s) - \iint g(t, s) d\Lambda_1(t) d\Lambda_2(s) \right. \\ &\quad \left. + \iint g(t, s) d\Lambda_1(t) d\hat{\Lambda}_2(s) - \iint g(t, s) d\Lambda_1(t) d\hat{\Lambda}_2(s) \right) \quad (4.105) \end{aligned}$$

add, subtract term

$$\begin{aligned} &= \sqrt{\frac{m}{n+m}} \iint g(t, s) d\hat{\Lambda}_2(s) d\sqrt{n} (\hat{\Lambda}_1(t) - \Lambda_1(t)) \\ &\quad + \sqrt{\frac{n}{n+m}} \iint g(t, s) d\Lambda_1(t) d\sqrt{m} (\hat{\Lambda}_2(s) - \Lambda_2(s)) + o_p(1) \quad (4.106) \end{aligned}$$

see explanation in (4.120) to (4.122) below

$$\xrightarrow{d} \sqrt{\frac{m}{n+m}} \int f_1(t) dB_1(C_1(t)) + \sqrt{\frac{n}{n+m}} \int f_2(s) dB_2(C_2(s)) \quad (4.107)$$

as $\min(n, m) \rightarrow \infty$ (see proof in Lemma 1), where

$$\int g(t, s) d\widehat{\Lambda}_2(s) \xrightarrow{p} \int g(t, s) d\Lambda_2(s) = f_1(t) \quad (4.108)$$

$$\int g(t, s) d\Lambda_1(t) = f_2(s) \quad (4.109)$$

$$C_1(t) = \int_{\gamma=0}^t \frac{d\Lambda_1(\gamma)}{(1 - F_1(\gamma^-))(1 - G_1(\gamma^-))} \quad (4.110)$$

$$C_2(s) = \int_{\gamma=0}^s \frac{d\Lambda_2(\gamma)}{(1 - F_2(\gamma^-))(1 - G_2(\gamma^-))} \quad (4.111)$$

$$B_1 \text{ is a Brownian motion} \quad (4.112)$$

$$B_2 \text{ is a Brownian motion} \quad (4.113)$$

$$\sqrt{n}(\widehat{\Lambda}_1(t) - \Lambda_1(t)) \xrightarrow{d} B_1(C_1(t)) \text{ as in Kalbfleisch [13]} \quad (4.114)$$

$$\sqrt{m}(\widehat{\Lambda}_2(s) - \Lambda_2(s)) \xrightarrow{d} B_2(C_2(s)) \text{ as in Kalbfleisch [13]}. \quad (4.115)$$

Therefore,

$$\sqrt{\frac{nm}{n+m}}(\overline{H}_{..} - \theta) \xrightarrow{d} N(0, \sigma^2) \text{ as } \min(n, m) \rightarrow \infty, \text{ from (4.107)} \quad (4.116)$$

$$\begin{aligned} \text{where } \sigma^2 = & \alpha \int f_1^2(t) \frac{d\Lambda_1(t)}{(1 - F_1(t^-))(1 - G_1(t^-))} \\ & + (1 - \alpha) \int f_2^2(s) \frac{d\Lambda_2(s)}{(1 - F_2(s^-))(1 - G_2(s^-))} \end{aligned} \quad (4.117)$$

and α is defined as in (4.127) below.

Therefore,

$$\frac{(\overline{H}_{..} - \theta)}{\sqrt{D}} \xrightarrow{d} N\left(0, \frac{\sigma^2}{\xi}\right) \text{ as } \min(n, m) \rightarrow \infty, \text{ using (4.116)} \quad (4.118)$$

where ξ is a constant such that

$$\frac{nm}{n+m} D \xrightarrow{p} \xi, \text{ as } \min(n, m) \rightarrow \infty. \quad (4.119)$$

We demonstrate that (4.119) is true in (4.124) to (4.141) below.

In (4.106) above we have assumed the following to be true:

$$\sqrt{\frac{nm}{n+m}} \iint g(t, s) d[\widehat{\Lambda}_1(t) - \Lambda_1(t)] d[\widehat{\Lambda}_2(s) - \Lambda_2(s)] \quad (4.120)$$

$$= \sqrt{\frac{1}{n+m}} \iint g(t, s) d\sqrt{n}[\widehat{\Lambda}_1(t) - \Lambda_1(t)] d\sqrt{m}[\widehat{\Lambda}_2(s) - \Lambda_2(s)] \quad (4.121)$$

$$\xrightarrow{d} \sqrt{\frac{1}{n+m}} \iint g(t, s) dB_1(C_1(t)) dB_2(C_2(s)) \quad (4.122)$$

as $\min(n, m) \rightarrow \infty$

$$= o_p(1) \quad (4.123)$$

With reference to (4.119) above, we demonstrate that $\frac{nm}{n+m} D \xrightarrow{p} \xi$ as $\min(n, m) \rightarrow \infty$ as follows:

$$D = \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{(Rx_i)^2} \frac{dy_j}{Ry_j} \overline{H}_{i,j} + \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{Rx_i} \frac{dy_j}{(Ry_j)^2} \overline{H}_{i,j} \quad (4.124)$$

from (4.60)

$$\begin{aligned} \frac{nm}{n+m}D &= \frac{m}{n+m} \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{Rx_i} \frac{dy_j}{Ry_j} \bar{H}_i. \\ &+ \frac{n}{n+m} \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{Rx_i} \frac{dy_j}{Ry_j} \frac{m}{Ry_j} \bar{H}_{.j} \end{aligned} \quad (4.125)$$

$$\begin{aligned} &\xrightarrow{p} \alpha \iint g(t,s) f_1(t) \frac{1}{(1-F_1(t^-))(1-G_1(t^-))} d\Lambda_1(t) d\Lambda_2(s) \\ &+ (1-\alpha) \iint g(t,s) f_2(s) \frac{1}{(1-F_2(s^-))(1-G_2(s^-))} d\Lambda_1(t) d\Lambda_2(s) \end{aligned} \quad (4.126)$$

as $\min(n, m) \rightarrow \infty$, by law of large numbers for Nelson-Aalen estimators, where

$$\text{We assume that } \frac{m}{n+m} \xrightarrow{p} \alpha \text{ as } \min(n, m) \rightarrow \infty, \quad 1 \geq \alpha > 0 \quad (4.127)$$

$$\text{Therefore } \frac{n}{n+m} \xrightarrow{p} 1 - \alpha \text{ as } \min(n, m) \rightarrow \infty \quad (4.128)$$

$$H_{ij} \text{ is represented as } g(t, s) \text{ as in (4.99)} \quad (4.129)$$

$$\bar{H}_i \xrightarrow{p} f_1(t), \text{ where } f_1(t) \text{ is as in (4.108), per Lemma 1 in section 6.5} \quad (4.130)$$

$$\bar{H}_{.j} \xrightarrow{p} f_2(s), \text{ where } f_2(s) \text{ is as in (4.109), per Lemma 1 in section 6.5} \quad (4.131)$$

$$\frac{n}{Rx_i} \xrightarrow{p} \frac{1}{(1-F_1(x_i))(1-G_1(x_i))} \quad \text{per Lemma 2 in section 6.5} \quad (4.132)$$

$$\frac{m}{Ry_j} \xrightarrow{p} \frac{1}{(1-F_2(y_j))(1-G_2(y_j))} \quad \text{per Lemma 2 in section 6.5} \quad (4.133)$$

$$G_1(t) \text{ is the censoring distribution of sample 1} \quad (4.134)$$

$$G_2(s) \text{ is the censoring distribution of sample 2} \quad (4.135)$$

$$\frac{dx_i}{Rx_i} \text{ is represented as } d\hat{\Lambda}_1 \text{ as in (4.100)} \quad (4.136)$$

$$\frac{dy_j}{Ry_j} \text{ is represented as } d\hat{\Lambda}_2 \text{ as in (4.101)} \quad (4.137)$$

$$= \alpha \int f_1^2(t) \frac{d\Lambda_1(t)}{(1 - F_1(t))(1 - G_1(t))} \quad (4.138)$$

$$+ (1 - \alpha) \int f_2^2(s) \frac{d\Lambda_2(s)}{(1 - F_2(s))(1 - G_2(s))} \quad \text{from (4.108) and (4.109)}$$

$$= \sigma^2 \quad \text{from (4.117)} \quad (4.139)$$

Hence,

$$\frac{nm}{n+m} D \rightarrow \sigma^2 \quad \text{as } \min(n, m) \rightarrow \infty, \quad \text{from (4.123) to (4.138)} \quad (4.140)$$

Combining (4.118) and (4.139) gives,

$$\frac{(\bar{H}_{..} - \theta)}{\sqrt{D}} \xrightarrow{d} N(0, 1) \quad \text{as } \min(n, m) \rightarrow \infty \quad (4.141)$$

Squaring both sides of (4.140) gives,

$$\frac{(\bar{H}_{..} - \theta)^2}{D} \xrightarrow{d} \chi_{(1)}^2 \quad \text{as } \min(n, m) \rightarrow \infty \quad (4.142)$$

Finally, combining (4.95) and (4.141) gives

$$-2ELLR \xrightarrow{d} \chi_{(1)}^2 \quad \text{as } \min(n, m) \rightarrow \infty \quad (4.143)$$

This concludes the main portion of the proof. We can use (4.143) to calculate an approximate p-value to test H_o based on $-2ELLR$.

4.5 Validation of Taylor Expansions

In this section we justify the Taylor expansions used in section 4.2 by showing that $\frac{\hat{\lambda}}{Rx_i} \tilde{H}_i \xrightarrow{p} 0$ as $\min(n, m) \rightarrow \infty$, as in (4.36); and $\frac{\hat{\lambda}}{Ry_j} \tilde{H}_j \xrightarrow{p} 0$ as $\min(n, m) \rightarrow \infty$, as in (4.37).

We will assume that \tilde{H}_i , \tilde{H}_j , \bar{H}_i , and \bar{H}_j are all uniformly bounded. We will also assume that $\xi \geq \frac{n}{m} \geq \frac{1}{\xi}$ for some positive number ξ . To simplify the calculations below we will additionally assume, without loss of generality, that $\xi = 1$, so that $n = m$.

First we show that $\hat{\lambda}^*/n$ becomes small as $n \rightarrow \infty$, where $\hat{\lambda}^*/n$ is the root of $\bar{H}_i - \theta - \lambda D = 0$.

From (4.118) we know that $\frac{(\bar{H}_i - \theta)}{\sqrt{D}} \rightarrow N\left(0, \frac{\sigma^2}{\frac{nm}{n+m}D}\right)$ and from (4.139) we know that $\frac{nm}{n+m}D \rightarrow \sigma^2$ as $\min(n, m) \rightarrow \infty$. Without loss of generality let us assume that $n = m$ so that $\frac{nm}{n+m} = \frac{n}{2}$. Then $\frac{nm}{n+m}D = \frac{n}{2}D$ so that

$$\frac{n}{2}D \xrightarrow{p} \sigma^2 \quad \text{from (4.139)} \quad (4.144)$$

We will show that $\hat{\lambda}^*/n \in (-\epsilon, \epsilon)$ as $n \rightarrow \infty$ for any small $\epsilon > 0$.

Now,

$$\bar{H}_{..} - \theta - \lambda D = \sqrt{D} \left(\frac{\bar{H}_{..} - \theta}{\sqrt{D}} - \lambda \sqrt{D} \right) \quad (4.145)$$

$$\doteq \sqrt{D} (Z - \lambda \sqrt{D}) \quad \text{where } Z \sim N(0, 1) \text{ from (4.140)} \quad (4.146)$$

$$= \sqrt{D} \left(Z - \lambda \sqrt{2/n} \sqrt{\frac{n}{2} D} \right) \quad (4.147)$$

$$\doteq \sqrt{D} \left(Z - \frac{\lambda}{\sqrt{n}} \sqrt{2} \sigma \right) \quad \text{from (4.143)} \quad (4.148)$$

Now we must choose a sequence in n that goes slowly to ∞ . Without loss of generality let us choose the function $\log(n)$. Then let $\frac{\lambda}{\sqrt{n}} = \pm \log n$ and substitute into (4.147) above. When $\frac{\lambda}{\sqrt{n}} = \log n$ then the RHS of (4.147) is less than 0 for large n , since Z is bounded in probability. And similarly when $\frac{\lambda}{\sqrt{n}} = -\log n$ then the RHS of (4.147) is greater than 0 for large n . This implies that $\frac{\hat{\lambda}^*}{n} \in \left(-\frac{\log n}{\sqrt{n}}, \frac{\log n}{\sqrt{n}} \right)$ for large n with probability approaching 1, since $\bar{H}_{..} - \theta - \lambda D$ is monotone in λ . This in turn implies that

$$\frac{\hat{\lambda}^*}{n} = O_p\left(\frac{\log n}{\sqrt{n}}\right) \quad (4.149)$$

Now,

$$\lim_{n \rightarrow \infty} \left(\frac{Rx_i}{n} \right) = (1 - F_1(x_i))(1 - G_1(x_i)) \quad \text{per Lemma 2, section 6.5} \quad (4.150)$$

which implies that

$$Rx_i = O_p(n) \quad (4.151)$$

Combining (4.148) and (4.152) gives

$$\frac{\widehat{\lambda}^*}{Rx_i} = O_p\left(\frac{\log n}{\sqrt{n}}\right) \quad (4.152)$$

Then since $\widetilde{H}_{i.}$ and $\widetilde{H}_{.j}$ are bounded, (4.152) implies

$$\frac{\widehat{\lambda}^*}{Rx_i} \widetilde{H}_{i.} = O_p\left(\frac{\log n}{\sqrt{n}}\right) \quad \text{and} \quad \frac{\widehat{\lambda}^*}{Ry_j} \widetilde{H}_{.j} = O_p\left(\frac{\log n}{\sqrt{n}}\right) \quad (4.153)$$

Then with reference to (4.44) (up to order $\widehat{\lambda}$) consider the following:

$$\begin{aligned} & \widehat{\lambda} \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{Rx_i} \frac{dy_j}{Ry_j} \left(\frac{\widetilde{H}_{i.}}{Rx_i} + \frac{\widetilde{H}_{.j}}{Ry_j} \right) \\ &= \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{Rx_i} \frac{dy_j}{Ry_j} \left(\frac{\widehat{\lambda}}{Rx_i} \widetilde{H}_{i.} + \frac{\widehat{\lambda}}{Ry_j} \widetilde{H}_{.j} \right) \end{aligned} \quad (4.154)$$

$$= O_p\left(\frac{\log n}{\sqrt{n}}\right) \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{Rx_i} \frac{dy_j}{Ry_j} \quad \text{from (4.152)} \quad (4.155)$$

$$= O_p\left(\frac{\log n}{\sqrt{n}}\right) \quad \text{using (4.45), since } \overline{H}_{..} \text{ is uniformly bounded} \quad (4.156)$$

Also with reference to (4.44) (up to order $\widehat{\lambda}^2$) consider the following:

$$\begin{aligned} & \widehat{\lambda}^2 \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{Rx_i} \frac{dy_j}{Ry_j} \left[\left(\frac{\widetilde{H}_{i.}}{Rx_i} \right)^2 + \frac{\widetilde{H}_{i.} \widetilde{H}_{.j}}{Rx_i Ry_j} + \left(\frac{\widetilde{H}_{.j}}{Ry_j} \right)^2 \right] \\ &= \sum_{i=1}^n \sum_{j=1}^m H_{ij} \frac{dx_i}{Rx_i} \frac{dy_j}{Ry_j} \left[\left(\frac{\widehat{\lambda} \widetilde{H}_{i.}}{Rx_i} \right)^2 + \frac{\widehat{\lambda}}{Rx_i} \widetilde{H}_{i.} \frac{\widehat{\lambda}}{Ry_j} \widetilde{H}_{.j} + \left(\frac{\widehat{\lambda} \widetilde{H}_{.j}}{Ry_j} \right)^2 \right] \end{aligned} \quad (4.157)$$

$$= O_p\left(\frac{\log n}{\sqrt{n}}\right)^2 \quad \text{from (4.152)} \quad (4.158)$$

Finally consider the following:

$$\tilde{H}_i = \bar{H}_i + (\tilde{H}_i - \bar{H}_i) \quad (4.159)$$

$$= \bar{H}_i + \sum_{j=1}^m H_{ij} \left(\hat{\nu}_j - \frac{dy_j}{Ry_j} \right) \quad \text{substituting (4.24) and (4.47)} \quad (4.160)$$

$$\doteq \bar{H}_i + \sum_{j=1}^m H_{ij} \left[\frac{dy_j}{Ry_j} \left(1 - \frac{\hat{\lambda}}{Ry_j} \tilde{H}_j \right) - \frac{dy_j}{Ry_j} \right] \quad \text{from (4.34) to order } \hat{\lambda} \quad (4.161)$$

$$= \bar{H}_i - \sum_{j=1}^m H_{ij} \frac{dy_j}{Ry_j} \left(\frac{\hat{\lambda}}{Ry_j} \tilde{H}_j \right) \quad \text{rearranging (4.160)} \quad (4.162)$$

$$= \bar{H}_i - \sum_{j=1}^m H_{ij} \frac{dy_j}{Ry_j} O_p \left(\frac{\log n}{\sqrt{n}} \right) \quad \text{using (4.152)} \quad (4.163)$$

$$= \bar{H}_i - O_p \left(\frac{\log n}{\sqrt{n}} \right) \sum_{j=1}^m H_{ij} \frac{dy_j}{Ry_j} \quad \text{rearranging (4.162)} \quad (4.164)$$

$$= \bar{H}_i - O_p \left(\frac{\log n}{\sqrt{n}} \right) \bar{H}_i. \quad \text{from (4.47)} \quad (4.165)$$

$$\doteq \bar{H}_i. \quad (4.166)$$

Therefore we conclude that the Taylor expansion in (4.47) is valid.

4.6 Lemmas

Lemma 1: For fixed t , $\int g(t, s) d\widehat{\Lambda}_2(s) \xrightarrow{p} \int g(t, s) d\Lambda_2(s)$. (4.167)

We will assume that $\int g(t, s) d\Lambda_2(s) < \infty$. We first apply the law of large numbers (LLN) to the case where the integral in (4.167) runs from 0 up to any large (but finite) positive number τ .

$$\int_0^\tau g(t, s) d\widehat{\Lambda}_2(s) = \sum_{j=1}^m I[y_j < \tau] g(t, y_j) \frac{dy_j}{Ry_j} \quad (4.168)$$

$$\begin{aligned} &= \frac{1}{m} \sum_{j=1}^m I[y_j < \tau] dy_j \left(\frac{g(t, y_j)}{Ry_j/m} - \frac{g(t, y_j)}{(1 - F_2(y_j^-))(1 - G_2(y_j^-))} \right) \\ &+ \frac{1}{m} \sum_{j=1}^m I[y_j < \tau] \frac{g(t, y_j) dy_j}{(1 - F_2(y_j^-))(1 - G_2(y_j^-))} \end{aligned} \quad (4.169)$$

The first term in (4.169) is bounded by

$$\frac{1}{m} \sum_{j=1}^m I[y_j < \tau] dy_j \left| \frac{g(t, y_j)}{Ry_j/m} - \frac{g(t, y_j)}{(1 - F_2(y_j^-))(1 - G_2(y_j^-))} \right| \quad (4.170)$$

$$\leq \sup_{y < \tau} \left| \frac{g(t, y)}{Ry/m} - \frac{g(t, y)}{(1 - F_2(y^-))(1 - G_2(y^-))} \right| \quad (4.171)$$

$$= \sup_{y < \tau} |g(t, y)| \left| \frac{1}{Ry/m} - \frac{1}{(1 - F_2(y^-))(1 - G_2(y^-))} \right| \quad (4.172)$$

Now $\sup_{y < \tau} |g(t, y)|$ is bounded by assumption. And in Lemma 2 below we show that $Ry_j/m \xrightarrow{p} (1 - F_2(y_j^-))(1 - G_2(y_j^-))$, hence its reciprocal is at least uniformly convergent on the set $\{y \leq \tau\}$. Hence the entire term in (4.172) is uniformly convergent to 0.

The last term in (4.169) is the average of an independent and identically distributed (iid) sum, so it converges to its expectation by the LLN. Now,

$$E\left(I[y_j < \tau] \frac{g(t, y_j) dy_j}{(1 - F_2(y_j^-))(1 - G_2(y_j^-))}\right) = \int_0^\tau g(t, s) d\Lambda_2(s) \quad (4.173)$$

Therefore, for any finite τ ,

$$\int_0^\tau g(t, s) d\widehat{\Lambda}_2(s) \xrightarrow{p} \int_0^\tau g(t, s) d\Lambda_2(s) \quad (4.174)$$

Now we address the tail $[\tau, \infty)$ in (4.167).

Consider the ratio,

$$\sup_{y_j \in [0, \infty)} \left| \frac{(1 - F_2(y_j^-))(1 - G_2(y_j^-))}{Ry_j/m} \right| \quad (4.175)$$

This ratio in (4.175) is bounded in probability (per Zhou(1991)[33]). Therefore, except on a set of probability η , for any $1 > \eta > 0$, we have,

$$\left| \frac{(1 - F_2(y_j^-))(1 - G_2(y_j^-))}{Ry_j/m} \right| < C, \quad \text{for some positive constant } C \quad (4.176)$$

Therefore,

$$\begin{aligned} & \sum_{j=1}^m I[y_j < \tau] |g(t, y_j)| \frac{dy_j}{Ry_j} \\ & \leq C \frac{1}{m} \sum_{j=1}^m I[y_j < \tau] \frac{|g(t, y_j)| dy_j}{(1 - F_2(y_j^-))(1 - G_2(y_j^-))} \end{aligned} \quad (4.177)$$

The RHS of (4.177) is an iid sum. Therefore by the strong law of large numbers it converges to its mean, which is,

$$C \int_\tau^\infty |g(t, s)| d\Lambda_2(s) \quad (4.178)$$

Since $\int |g(t, s)|d\Lambda_2(s) < \infty$ by assumption, therefore $\int_{\tau}^{\infty} |g(t, s)|d\Lambda_2(s)$ in (4.178) can be made arbitrarily small by selecting a large value of τ , such that $\int g(t, s)d\Lambda_2(s) < \frac{\epsilon}{C}$, where ϵ is an arbitrarily small positive number. Therefore,

$$C \int_{\tau}^{\infty} g(t, x)d\Lambda_2(s) < \epsilon \quad (4.179)$$

$$\text{Thus} \quad \int_{\tau}^{\infty} g(t, s)d\widehat{\Lambda}_2(s) \xrightarrow{p} 0 \quad (4.180)$$

$$\text{and} \quad \int_0^{\infty} g(t, s)d\widehat{\Lambda}_2(s) \xrightarrow{p} \int_0^{\infty} g(t, s)d\Lambda_2(s) \quad (4.181)$$

Lemma 2: $\frac{m}{Ry_j} \xrightarrow{p} \frac{1}{(1-F_2(y_j))(1-G_2(y_j))}$

$$\lim_{m \rightarrow \infty} \left(\frac{Ry_j}{m} \right) = \lim_{m \rightarrow \infty} \left(\frac{\sum_{r=1}^m I[y_r \geq y_j]}{m} \right) \quad (4.182)$$

$$= \mathbb{E}(I[y_r \geq y_j]) \quad \text{Glivenko-Cantelli, law of large numbers} \quad (4.183)$$

$$= P(y_r \geq y_j) \quad (4.184)$$

$$= (1 - F_1(y_j))(1 - G_1(y_j)) \quad (4.185)$$

Equation (4.169) involves the variable Y as well as the right-censoring variable (which we may call R) associated with Y . We only observe the variable $T = \min(Y, R)$. In order for T to exceed a value y_j it is necessary that both Y and R exceed y_j . Hence the probability that T exceeds y_j is the product of the probability that Y exceeds y_j and the probability that R exceeds y_j , where a product applies since Y and R are assumed to be independent.

Then,

$$\lim_{m \rightarrow \infty} \left(\frac{m}{Rx_i} \right) = \lim_{m \rightarrow \infty} \left(\frac{1}{Ry_j/m} \right) \quad (4.186)$$

$$= \frac{1}{\lim_{m \rightarrow \infty} (Ry_j/m)} \quad \text{since } s(t) = 1/t \text{ is continuous for } t \neq 0 \quad (4.187)$$

$$= \frac{1}{(1 - F_2(y_j))(1 - G_2(y_j))} \quad (4.188)$$

Similarly, regarding (4.132) in Part 3 above,

$$\lim_{n \rightarrow \infty} \left(\frac{n}{Rx_i} \right) = \frac{1}{(1 - F_1(x_i))(1 - G_1(x_i))} \quad (4.189)$$

Chapter 5 Relationship between Constrained Kaplan-Meier Estimators and the Corresponding Nelson-Aalen Estimators

5.1 Background

In this chapter we study the relationship between the constrained Kaplan-Meier estimator and the corresponding constrained Nelson-Aalen estimator, for right-censored data.

Akritis (2000) [1] explores the relationship between the Kaplan-Meier estimator and the corresponding Nelson-Aalen estimator, for right-censored data. We shall extend the work of Akritis by exploring the relationship between the constrained Kaplan-Meier estimator and the corresponding constrained Nelson-Aalen estimator, for right-censored data. We shall argue that once this relationship is established in a certain format, many distribution-type estimation problems can be converted to hazard-type estimation problems, which are more easily solved when the data is right-censored.

First let us discuss how we may establish the equivalency of an estimating equation in terms of hazard and an estimating equation in terms of distribution, for right-censored data. (See Zhou's private notes (2005) [35].)

Consider an estimating equation that utilizes a function g ,

$$0 = \frac{1}{n} \sum_{i=1}^n g(X_i, \theta).$$

We may rewrite this as

$$0 = \int g(x, \theta) d\hat{F}_n(x), \tag{5.1}$$

where $\hat{F}_n(\cdot)$ is the empirical distribution based on X_i 's. Such an estimating equation is typically based on the premise that the estimator is unbiased, so that

$$0 = \int g(x, \theta_0) dF(x),$$

where $F(x)$ is the true CDF of X_i , and θ_0 is the true parameter value. Then by the Central Limit Theorem (CLT) we know that

$$\sqrt{n} \left[\int g(x, \theta) d\hat{F}_n(x) - \int g(x, \theta_0) dF(x) \right]$$

converges to a normal distribution with zero mean.

The estimating equation in (5.1) above suggests a similar equation for right-censored data:

$$0 = \int g(x, \theta) d\hat{F}_{KM}(x),$$

where $\hat{F}_{KM}(\cdot)$ is the Kaplan-Meier estimator based on the censored sample (T_i, δ_i) . The assumption that the estimator is unbiased,

$$\int g(x, \theta_0) dF(x) = 0,$$

then leads to

$$0 = \int g(x, \theta) d[\hat{F}_{KM}(x) - F(x)].$$

Now suppose we could find a non-random function $g^*(t, \theta)$ such that

$$0 = \int g(x, \theta) d[\hat{F}_{KM}(x) - F(x)] \Leftrightarrow 0 = \int g^*(t, \theta) d[\hat{\Lambda}_{NA}(t) - \Lambda(t)], \quad (5.2)$$

where $\Lambda_{NA}(\cdot)$ is the cumulative hazard function based on \hat{F}_{KM} ,

$$\hat{\Lambda}_{NA}(t) = \int_{[0,t)} \frac{d\hat{F}_{KM}(s)}{1 - \hat{F}_{KM}(s-)},$$

and Λ is similarly based on F ,

$$\hat{\Lambda}(t) = \int_{[0,t)} \frac{dF(s)}{1 - F(s-)}.$$

Then it is well-known that $\hat{\Lambda}_{NA}(t)$ is just the Nelson-Aalen estimator, and the estimating equation in the LHS of (5.2) can be replaced by the estimating equation in terms of hazard in the RHS of (5.2).

Unfortunately, the g^* that satisfies the “if and only if” requirement part above is hard to find. However, we do not have to find an exact “if and only if” substitute, rather an asymptotic equivalency can be established if the two are within an $o_p(1/\sqrt{n})$ of each other. Alternatively, an approximate equivalency can be established if we can find a θ which makes the two both close to zero.

Along these lines, Akritas (2000) established the following equivalency in his Proposition 3 under some regularity conditions:

$$\int g(s, \theta) d[\hat{F}_{KM}(s) - F(s)] = \int \tilde{g}(s, \theta) d[\hat{\Lambda}_{NA}(s) - \Lambda s] + o_p(n^{-1/2}) \quad (5.3)$$

where \tilde{g} is non-random and given in his equation (9).

To use this type of equivalency in the empirical likelihood analysis we need to establish this relationship not only for the Kaplan-Meier estimator and Nelson-Aalen estimator, but also for other estimators as well. The reason is that the likelihood ratio involves two likelihoods: one involves Kaplan-Meier/Nelson-Aalen in the denominator, the other has to do with maximization under constraints and are not achieved by the Kaplan-Meier/Nelson-Aalen estimator.

We will lay a basis for the following relationship under certain conditions in the next section:

For any discrete CDF F^* that is dominated by the Kaplan-Meier estimator and

is close to the Kaplan-Meier estimator, that is,

$$F^* \ll \hat{F}_{KM} \quad \text{and} \quad \|F^* - \hat{F}_{KM}\| = O(1/\sqrt{n}) \quad (5.4)$$

the above equivalency relationship (5.3) still holds when \hat{F}_{KM} is replaced by F^* and the Nelson-Aalen estimator $\hat{\Lambda}_{NA}$ is replaced by $\Lambda^*(t)$, where

$$\Lambda^*(t) = \int_{[0,t)} \frac{dF^*(s)}{1 - F^*(s-)}.$$

5.2 Relationship Between the Constrained Kaplan-Meier Estimator and the Corresponding Nelson-Aalen Estimator

The data setting as described below is the same as Akritas (2000).

Let $T_i, i = 1, \dots, n$, be i.i.d. random variables on the real line and let F denote their common distribution function. Let the survival function of the T_i be denoted by S , and the cumulative hazard function by Λ . Thus $S = 1 - F$ and $\Lambda(t) = \int_{-\infty}^t [S(x-)]^{-1} dF(x)$. The observed data consist of

$$X_i = \min(T_i, C_i) \quad \text{and} \quad \Delta_i = I(X_i = T_i), \quad i = 1, \dots, n, \quad (5.5)$$

where C_1, \dots, C_n are i.i.d. random censoring variables which are independent of the T_i . The common distribution function of the C_i is denoted by G . The distribution function of the X_i is denoted by H . Thus, $1 - H = (1 - F)(1 - G)$. $I(E)$ denotes the indicator of the event E .

Let $N_i(t) = I(X_i \leq t, \Delta_i = 1)$, $N \cdot(t) = \sum_{i=1}^n N_i(t)$, $Y_i(t) = I(X_i \geq t)$, $Y \cdot(t) = \sum_{i=1}^n Y_i(t)$. We will assume that all random variables are defined on the probability space (Ω, \mathcal{F}, P) and we will consider the filtration

$$\mathcal{F}_t = \mathcal{N} \bigvee \sigma\{(X_i, \Delta_i)I(X_i \leq s), I(X_i > s) : -\infty < s \leq t, i = 1, \dots, n\}, \quad (5.6)$$

where \mathcal{N} consists of all P -null sets of \mathcal{F} . Then, $M_i(t) = N_i(t) - \int_{-\infty}^t Y_i(s)d\Lambda(s)$ is a martingale with respect to the filtration in (5.6). Similarly, we denote $M.(t) = \sum_{i=1}^n M_i(t)$. The Kaplan-Meier estimator of S based on the observations (5.5) will be denoted by \hat{S} , while $\hat{F} = 1 - \hat{S}$ and $\hat{\Lambda}$ will denote the corresponding estimators of F and Λ . The constrained Kaplan-Meier estimator of S will be denoted by S^* , while $F^* = 1 - S^*$ and Λ^* will denote the corresponding constrained estimators of F and Λ . $A_-(s)$ or $A(s-)$ will denote the left-continuous version of a right-continuous function A , and we define $\Delta A(s) = A(s) - A(s-)$. Unless otherwise explicitly indicated, the domain of integration includes the upper and lower integration limits. Finally, define $\tau_n = \max(X_1, \dots, X_n)$, $\tau = \tau_H = \inf\{x : H(x) = 1\} \leq \infty$, and let $\tau_F = \sup\{x : F(x) < 1\}$, for any distribution function F . We also assume $\phi : \mathfrak{R} \rightarrow \mathfrak{R}$ is any measurable function such that $\int \phi^2 dF < \infty$.

Assumption 1. (Akritas (2000)) *Let $\tau = \tau_H$. Then*

$$\int_{-\infty}^{\tau} \frac{\phi^2(s)}{1 - G(s-)} dF(s) < \infty.$$

Assumption 1 implies the following inequality, which we will use later on:

$$\int_{-\infty}^{\tau} \phi^2(s) dF(s) < \infty.$$

Assumption 2. *Let F^* be defined as above. Then*

$$F^* \ll \hat{F}_{KM} \quad \text{and} \quad \|F^* - \hat{F}_{KM}\| = O(1/\sqrt{n}).$$

Proposition 1. *Under Assumption 1 and 2 there exists a sequence of constants $K_n \rightarrow \infty$ such that the function ϕ truncated at K_n , $\phi_n(s) = \phi(s)I(|\phi(s)| \leq K_n)$, satisfies*

$$\int_{(-\infty, \tau]} \phi(s) d(F^*(s) - F(s)) = \int_{(-\infty, \tau]} \phi_n(s) d(F^*(s) - F(s)) + o_p(n^{-1/2}).$$

Proof. First, it will be shown that for any sequence K_n such that $K_n n^{-1/2} \rightarrow \infty$,

$$\int_{-\infty}^{\tau} \phi(s) I(|\phi(s)| > K_n) dF(s) = o_p(n^{-1/2}). \quad (5.7)$$

This is the same as Akritas (2000). Set $\phi(s) = 0$ for $s > \tau$, and write

$$\begin{aligned} \int_{-\infty}^{\tau} \phi(s) I(|\phi(s)| > K_n) dF(s) &= \int_{-\infty}^{\infty} x I(|x| > K_n) dF_{\phi}(x) \\ &= K_n(1 - F_{\phi}(K_n)) + \int_{K_n}^{\infty} (1 - F_{\phi}(x)) dx - K_n F_{\phi}(-K_n) \\ &\quad - \int_{-\infty}^{-K_n} F_{\phi}(x) dx, \end{aligned} \quad (5.8)$$

where $F_{\phi}(x) = \int I(\phi(s) \leq x) dF(s)$.

We will show that the first and second terms on the right-hand side of (5.8) are $o_p(n^{-1/2})$; similar arguments apply for the other two terms. For the first term, write (for large n)

$$\sqrt{n} K_n (1 - F_{\phi}(K_n)) \leq K_n^2 (1 - F_{\phi}(K_n)) \leq \int_{K_n}^{\infty} x^2 dF_{\phi}(x) \rightarrow 0$$

as $n \rightarrow \infty$, where we use $K_n n^{-1/2} \rightarrow \infty$ for the first inequality and Assumption 1 for the convergence to zero. Replacing K_n by x in the middle inequality, we obtain

$$x^2 (1 - F_{\phi}(x)) \rightarrow 0,$$

as $x \rightarrow \infty$, implying that, for large x ,

$$1 - F_{\phi}(x) \leq x^{-2}.$$

From this it follows that

$$\sqrt{n} \int_{K_n}^{\infty} (1 - F_{\phi}(x)) dx \leq \sqrt{n} \int_{K_n}^{\infty} x^{-2} dx = \sqrt{n} K_n^{-1} \rightarrow 0.$$

Thus (5.8) is proven, and so, consequently, is (5.7). Next, it will be shown that there is a sequence $K_n \rightarrow \infty$ such that

$$\int_{-\infty}^{\tau} \phi(s) I(|\phi(s)| > K_n) dF^*(s) = o_p(n^{-1/2}). \quad (5.9)$$

Let $X_{(n_2)}$ denote the largest uncensored observation. Then we can write

$$\begin{aligned} & |\sqrt{n} \int_{-\infty}^{\tau} \phi(s) I(\phi(s) > K_n) dF^*(s)| \\ & \leq \max_{1 \leq i \leq n} \{|\phi(X_i) \Delta_i|\} \sqrt{n} \int_{-\infty}^{\tau} I(|\phi(s)| > K_n) dF^*(s) \\ & \leq \max_{1 \leq i \leq n} \{|\phi(X_i) \Delta_i|\} \sqrt{n} \sum_{i=1}^n I(|\phi(X_i) \Delta_i| > K_n) (\Delta \hat{F}(X_{(n_2)}) + \frac{C}{\sqrt{n}}) \\ & = \max_{1 \leq i \leq n} \{|\phi(X_i) \Delta_i|\} \sqrt{n} \sum_{i=1}^n I(|\phi(X_i) \Delta_i| > K_n) (1 - \hat{F}(X_{(n_2)})) \frac{\Delta N \cdot (X_{(n_2)})}{Y \cdot (X_{(n_2)})} \\ & + C \max_{1 \leq i \leq n} \{|\phi(X_i) \Delta_i|\} \sum_{i=1}^n I(|\phi(X_i) \Delta_i| > K_n) \\ & \leq B \max_{1 \leq i \leq n} \{|\phi(X_i) \Delta_i|\} \sqrt{n} \sum_{i=1}^n I(|\phi(X_i) \Delta_i| > K_n) \frac{1 - F(X_{(n_2)} -)}{1 - H(X_{(n_2)} -)} \frac{\Delta N \cdot (X_{(n_2)})}{n} \quad (5.10) \\ & + C \max_{1 \leq i \leq n} \{|\phi(X_i) \Delta_i|\} \sum_{i=1}^n I(|\phi(X_i) \Delta_i| > K_n) \end{aligned}$$

The second inequality holds because $\|F^* - \hat{F}_{KM}\| = O(1/\sqrt{n})$, where C is a constant. The positive constant B in the first term of the last inequality can be chosen large enough by Lemmas 2.6 and 2.7 of Gill (1983) for the last inequality in (5.10) to hold with probability as high as desired. Using Markov's inequality and

Assumption 1, it can be seen that, for any sequence $K_n \rightarrow \infty$,

$$R_{1,n} = \frac{K_n^2}{n} \sum_{i=1}^n I(|\phi(X_i)\Delta_i| > K_n) = o_p(1) \quad (5.11)$$

$$R_{2,n} = n^{-1/2} \max_{1 \leq i \leq n} \{|\phi(X_i)\Delta_i|\} = o_p(1) \quad (5.12)$$

Combining (5.10), (5.11) and (5.12) we can write

$$\begin{aligned} \sqrt{n} \int_{-\infty}^{\tau} \phi(s) I(|\phi(s)| > K_n) dF^*(s) &\leq BR_{1,n} R_{2,n} \frac{n^2}{K_n^2} \frac{1 - F(X_{(n_2)-})}{1 - H(X_{(n_2)-})} \frac{\Delta N \cdot (X_{(n_2)})}{n} \\ &+ CR_{1,n} R_{2,n} \frac{n^2}{K_n^2} \frac{1}{\sqrt{n}} \end{aligned} \quad (5.13)$$

with the inequality holding for n large enough and with probability as high as desired by choosing B large enough. Akritas gives a detailed proof that the first term in the right-hand side of (5.13) is $o_p(1)$. And it is evident that the second term is also $o_p(1)$ by choosing $K_n n^{-1} \rightarrow \infty$. Therefore the left-hand side in (5.13) is $o_p(1)$, which implies that relation (5.9) holds, and hence Proposition 1 is proven. ■

Proposition 2. *Let $\phi_n(s)$ be as defined in Proposition 1. Then, under Assumption 1,*

$$\int_{(-\infty, \tau]} \phi_n(s) d(F^*(s) - F(s)) = \int_{(-\infty, \tau_n]} \phi_n(s) d(F^*(s) - F(s)) + o_p(n^{-1/2}).$$

Proof. See Akritas (2000).

The following lemma will be used to prove the theorem below.

Lemma. (Gill 1980) *Let A and B be right continuous nondecreasing functions on $[0, \infty)$, zero at time zero; suppose $\Delta A \leq 1$ and $\Delta B < 1$ on $[0, \infty)$. Then the unique locally bounded solution Z of*

$$Z(t) = \int_{s \in [0, t]} \frac{1 - Z(s-)}{1 - \Delta B(s)} (dA(s) - dB(s))$$

is given by

$$Z(t) = 1 - \frac{\prod_{s \leq t} (1 - \Delta A(s)) \exp(-A_c(t))}{(1 - \Delta B(s)) \exp(-B_c(t))}$$

where it should be recalled that A_c is the continuous parts of A , defined by

$$A_c(t) = A(t) - \sum_{s \leq t} \Delta A(s).$$

Theorem. (by Li Liu) Let $G = \int (1 - F_-)^{-1} dF$ for some sub-distribution F with $F(0) = 0$, and define $\tau = \sup\{t : F(t) < 1\}$.

1. $F(t) = \int_{s \in [0, t]} (1 - F(s-)) dG(s)$ uniquely determines F if G is given; and F can be written as

$$F(t) = 1 - \prod_{s \leq t} (1 - \Delta G(s)) \exp(-G_c(t)) \quad \text{for all } t.$$

2. F and G are constant on $[\tau, \infty)$, G is finite and $\Delta G < 1$ on $[0, \tau)$. If $F(\tau-) < 1$, then $G(\tau) < \infty$ and $\Delta G(\tau) = 1$ iff $F(\tau) = 1$. If on the other hand $F(\tau-) = 1$, then $G(t) \uparrow G(\tau) = \infty$ as $t \uparrow \tau$.

3. If F has a density f , then defining the hazard rate or failure rate λ by $\lambda = f/(1 - F)$,

$$G(t) = \int_{s \in [0, t]} \lambda(s) ds \quad \text{for all } t.$$

More generally, if F is continuous, we have

$$G = -\log(1 - F).$$

4. For all t such that $F(t) < 1$,

$$\frac{1 - \hat{F}(t)}{1 - F(t)} = 1 - \int_0^t \frac{1 - \hat{F}(s-)}{1 - F(s)} (d\hat{G}(s) - dG(s))$$

where \hat{F} (and \hat{G}) is just another function satisfying the same conditions as F .

Proposition 3. *Suppose that either $\tau_n < \tau_F$ a.s., or that $\phi(\tau_F) = 0$. Define*

$$\tilde{\phi}_n(s) \equiv \tilde{\phi}_n(s; \tau) = S(s-) \left[\phi_n(s) - \frac{1}{S(s)} \int_{(s, \tau]} \phi_n(t) dF(t) \right].$$

Then, under Assumption 1 and 2,

$$\int_{-\infty}^{\tau_n} \phi_n(s) d(F^*(s) - F(s)) = \int_{-\infty}^{\tau_n} \tilde{\phi}_n(s) d(\Lambda^*(s) - \Lambda(s)) + o_p(n^{-1/2}).$$

Proof. First, using the identity

$$\frac{1 - F^*(t)}{1 - F(t)} = 1 - \int_0^t \frac{1 - F^*(s-)}{1 - F(s)} d(\Lambda^*(x) - \Lambda(x))$$

(Li Liu (2008) [19]), or equivalently,

$$F^*(t) - F(t) = S(t) \int_{-\infty}^t S^*(x-) [S(x)]^{-1} d(\Lambda^*(x) - \Lambda(x))$$

and a straightforward calculation, it follows that

$$\int_{-\infty}^{\tau_n} \phi_n(s) d(F^*(s) - F(s)) = \int_{-\infty}^{\tau_n} \hat{\phi}_n^*(s; \tau_n) d(\Lambda^*(s) - \Lambda(s)) \quad (5.14)$$

where $\hat{\phi}_n^*(s; t) = S^*(s-) [\phi_n(s) - [S(s)]^{-1} \int_{(s, t]} \phi_n(x) dF(x)]$, for all t .

The rest of the proof needs further investigation.

Chapter 6 Simulations and Examples

6.1 Simulations

6.1.1 Simulation 1.

In this simulation, we show that the test statistic T has an asymptotic $\chi_{(1)}^2$ distribution under the null hypothesis. We generate the censored survival data from the following setting:

Survival Time Distribution:	$F_0(t) = 1 - e^{-0.02t}$
Censoring Distribution:	$G_0(t) = 1 - e^{-0.005t}$
Cumulative Hazard Function:	$\Lambda_0(t) = 0.02t$
Sample Size:	n=50
Constraint Equation 1:	$g_1(t) = I(t \leq 20)$
Constraint Equation 2:	$g_2(t) = I(20 < t \leq 40)$
Parameter θ_0 :	$\theta_0 = \int_0^\infty g_1(t)d\Lambda_0(t) = 0.4$
	$\theta_0 = \int_0^\infty g_2(t)d\Lambda_0(t) = 0.4$
Number of Simulations:	1000

The result is shown in Figure 6.1 . The horizontal axis are $\chi_{(1)}^2$ quantiles, and the vertical axis are sorted T test statistic. We can see that under the null hypothesis: $H_0 : \theta = 0.4$, the test statistic T has an asymptotic $\chi_{(1)}^2$ distribution.

If we increase the sample size to 100, we can see, from Figure 6.2, the test statistic T fits the asymptotic $\chi_{(1)}^2$ distribution better.

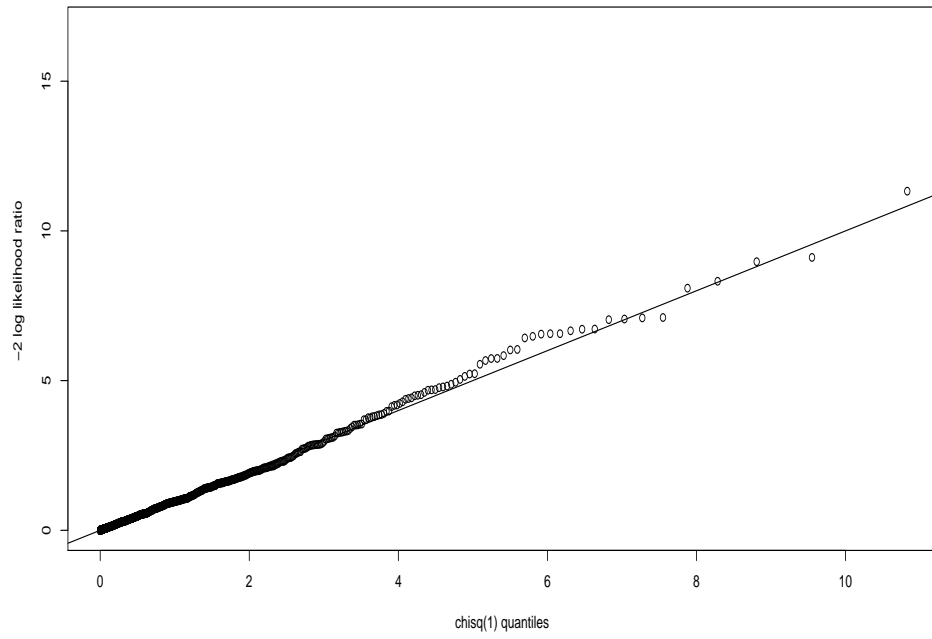


Figure 6.1: Simulation-sample size 50 with true parameter

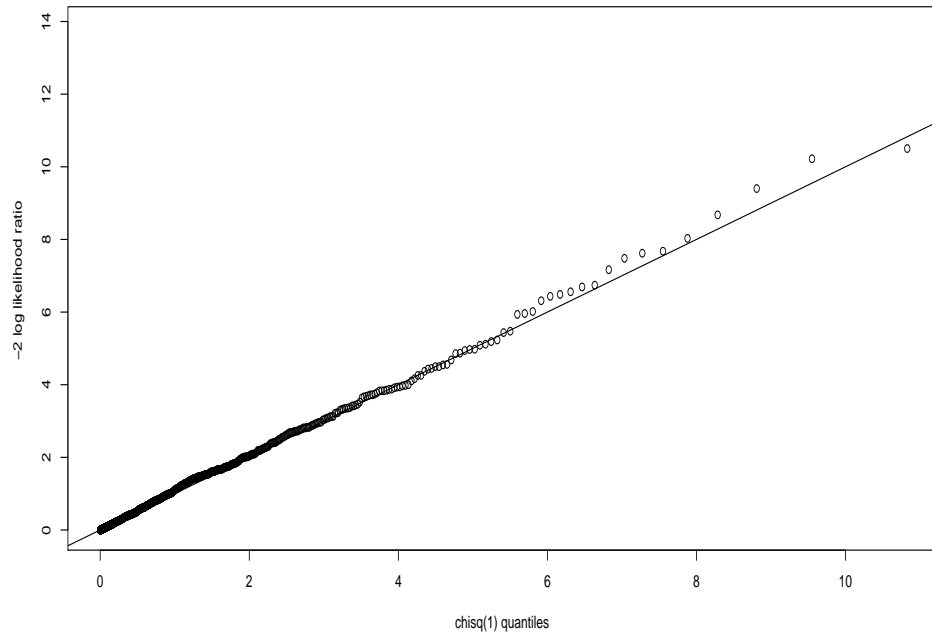


Figure 6.2: Simulation-sample size 100 with true parameter

Now we change the censoring distribution to be: 20+Exponential ($\lambda_1 = 0.02$), and the test statistic T still has an asymptotic $\chi^2_{(1)}$ distribution under the null hypothesis.

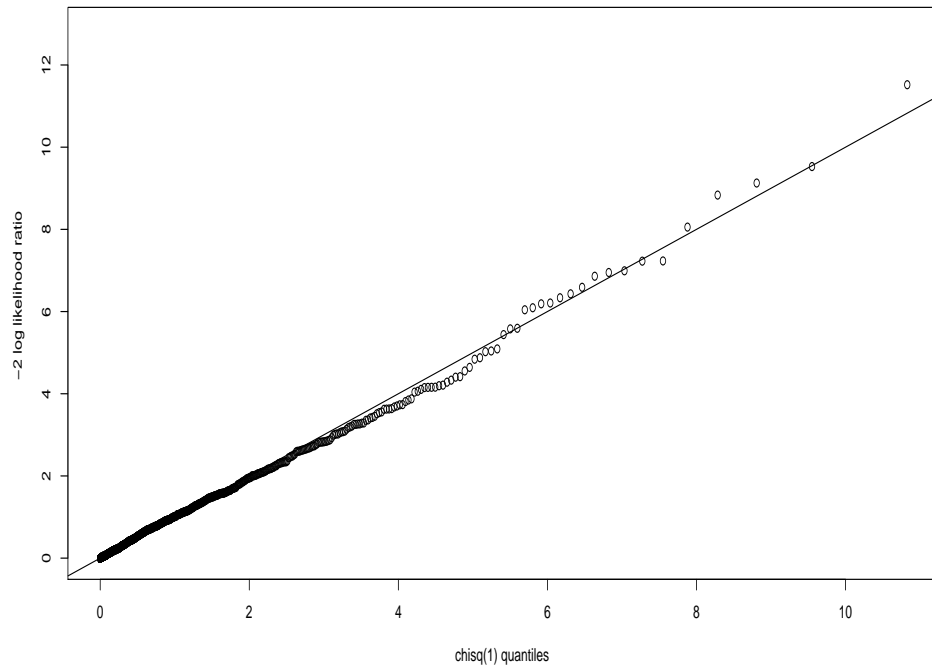


Figure 6.3: Simulation-with censoring distribution: 20+Exponential ($\lambda_1 = 0.02$)

Figure 6.4 simply shows that, under the alternative hypothesis, the distribution definitely is not $\chi^2_{(1)}$.

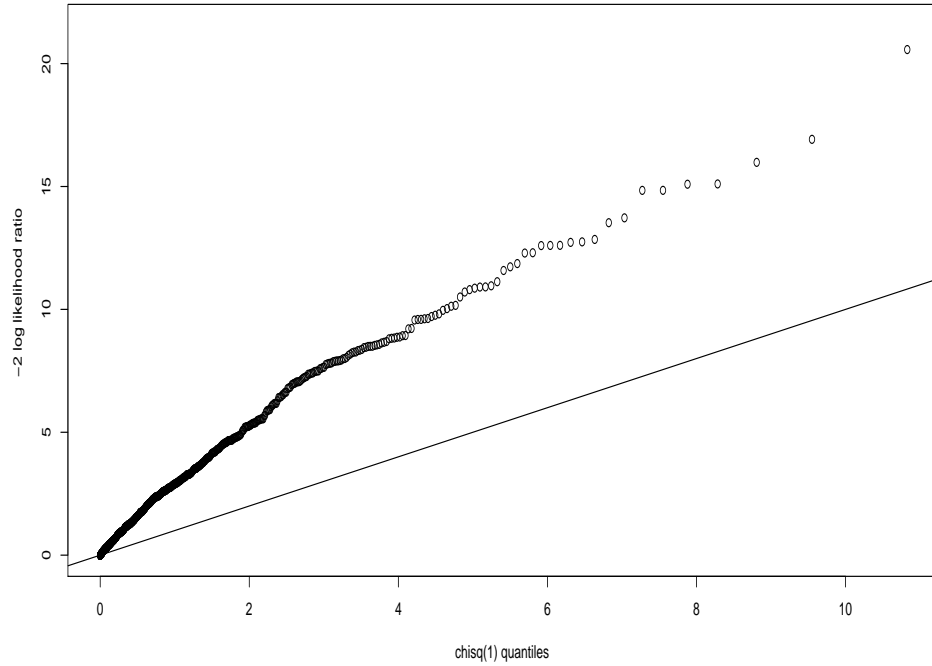


Figure 6.4: Simulation-sample size 50 with false parameter

Remark: Indicator function is a simple function. Although it does not have a derivative or not smooth, the mean-type integral we deal with is still valid and we still have the chi-square result.

6.1.2 Simulation 2.

We will show the efficiency of the maximum empirical likelihood estimator $\hat{\theta}$ obtained in Theorem 3 in this simulation. The censored survival data in the first 6 plots were generated from the same setting as in *Simulation 1*. The censoring distribution used in the rest 6 plots is: 20+Exponential ($\lambda_1 = 0.02$).

In the following figures, we calculate the empirical likelihood ratio statistic T with different constraints. The first plot was generated with only the first constraint equation, $g_1(t) = I(t \leq 20)$, for different θ s (just-determined case); the second plot used only the second constraint, $g_2(t) = I(20 < t \leq 40)$, to calculate the empirical likelihood ratio statistic T for different θ s (just-determined case); the last plot gives the empirical likelihood ratio statistic T with both constraints for different θ s (over-determined case).

It is obvious that the last plot has the sharpest empirical likelihood ratio curve. So we should expect a shorter 95% confidence interval from the over-determined case than from the just-determined case, or equivalently, the maximum empirical likelihood estimator $\hat{\theta}$ is more efficient for the over-determined case than for the just-determined case.

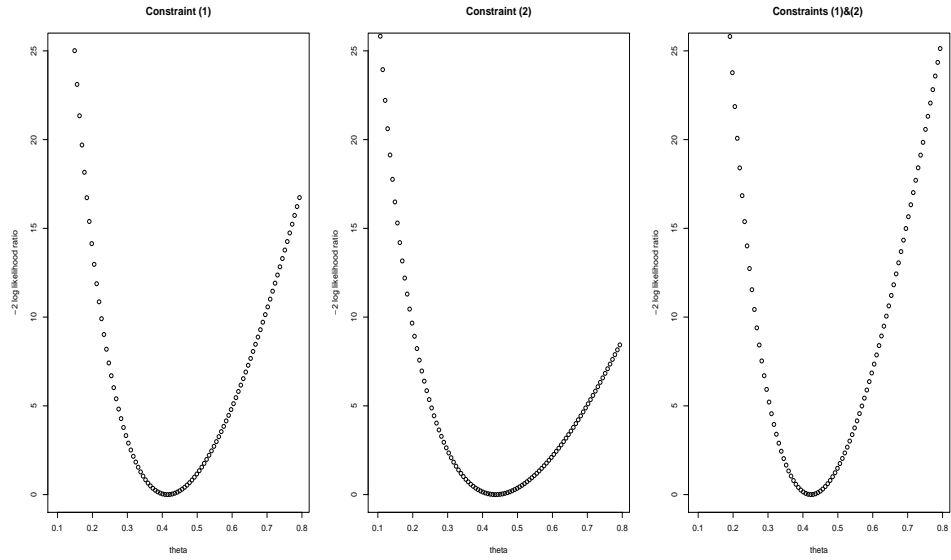


Figure 6.5: Empirical likelihood ratio statistic with different constraints, $n=100$

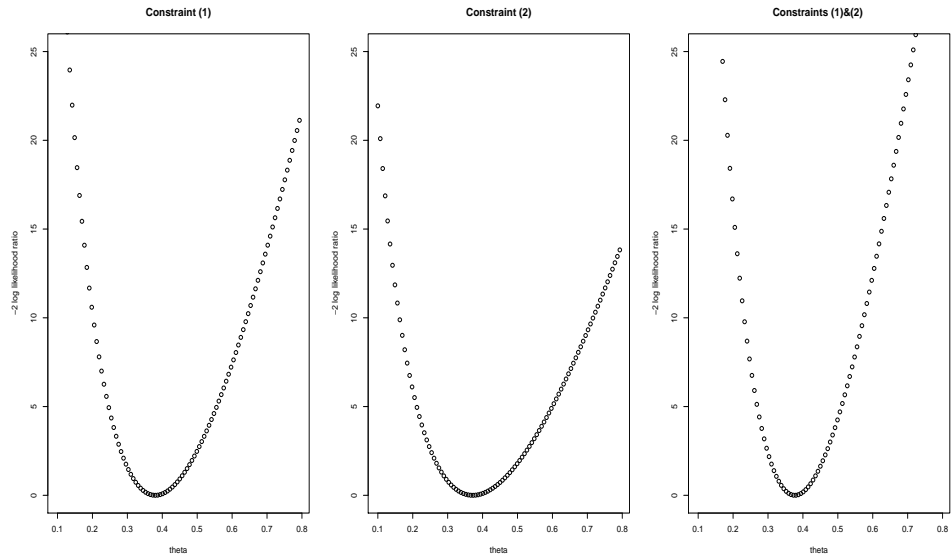


Figure 6.6: Empirical likelihood ratio statistic with different constraints, $n=100$

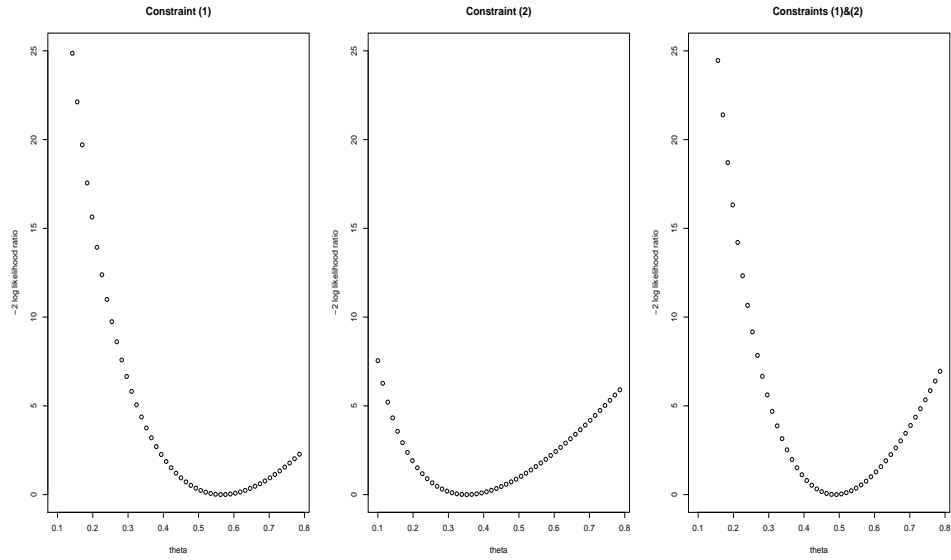


Figure 6.7: Empirical likelihood ratio statistic with different constraints, $n=50$

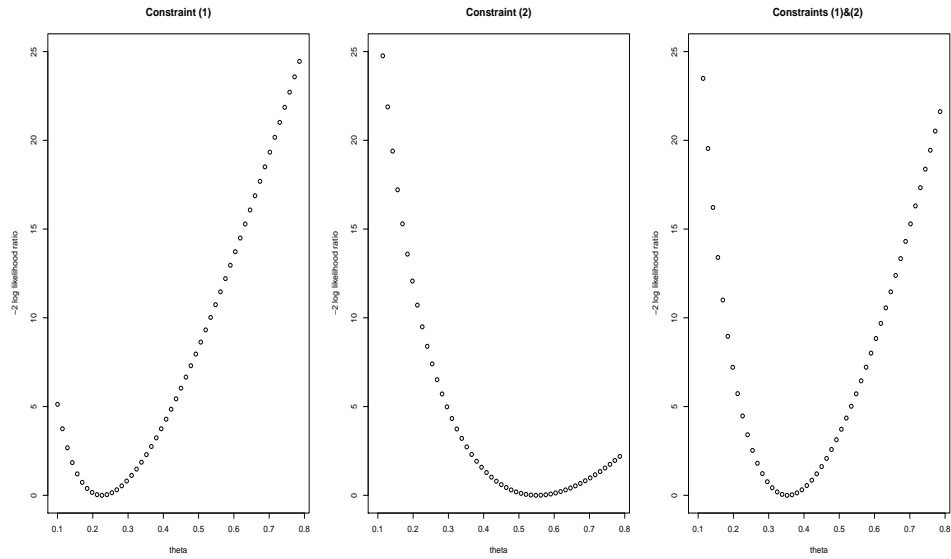


Figure 6.8: Empirical likelihood ratio statistic with different constraints, $n=50$

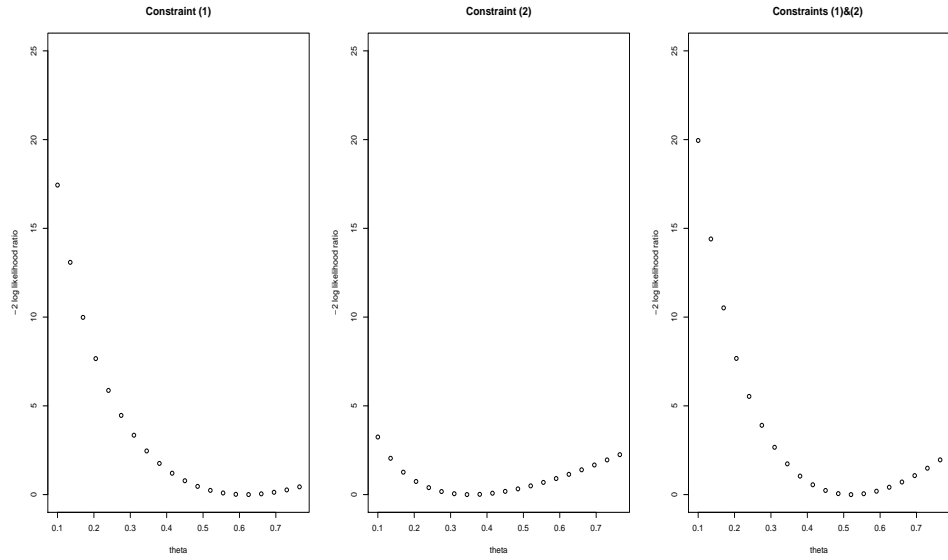


Figure 6.9: Empirical likelihood ratio statistic with different constraints, $n=20$

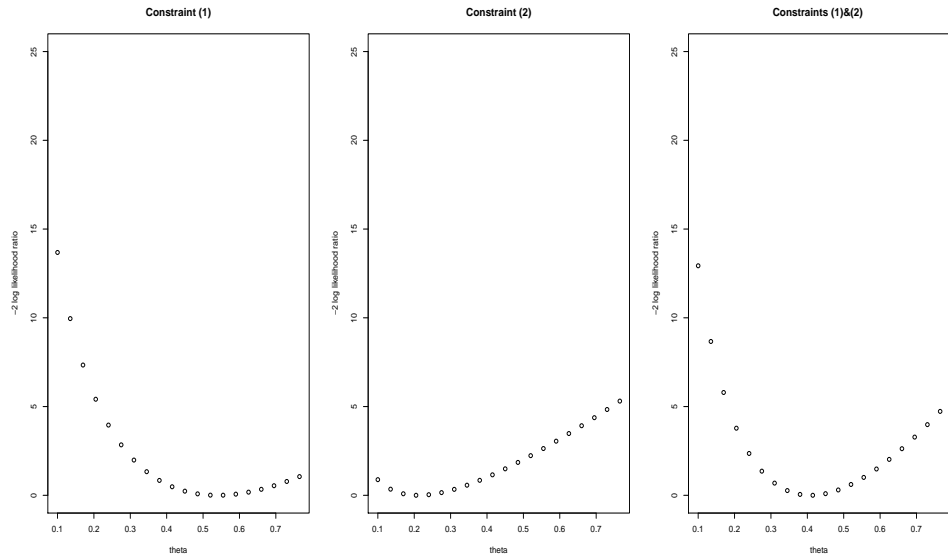


Figure 6.10: Empirical likelihood ratio statistic with different constraints, $n=20$

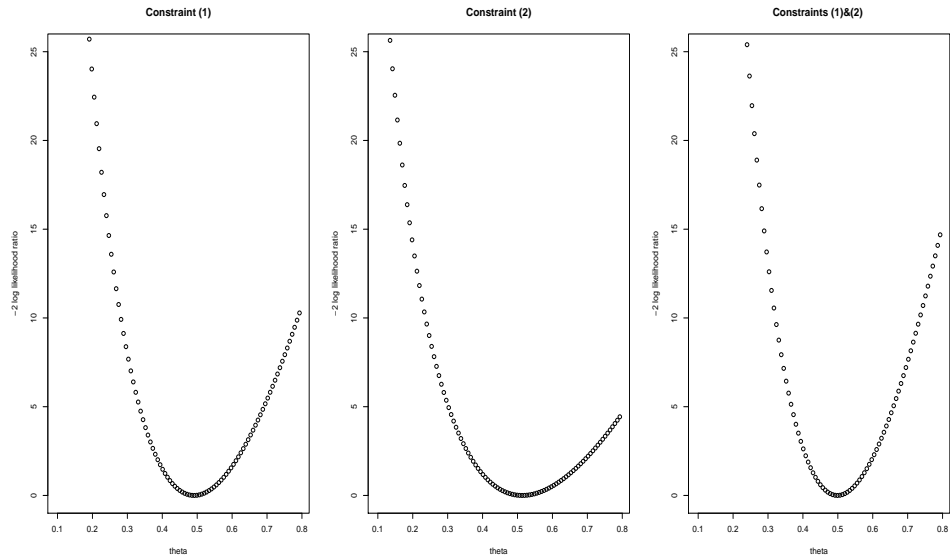


Figure 6.11: Empirical likelihood ratio statistic with different constraints and censoring distribution: 20+Exponential ($\lambda_1 = 0.02$), $n=100$

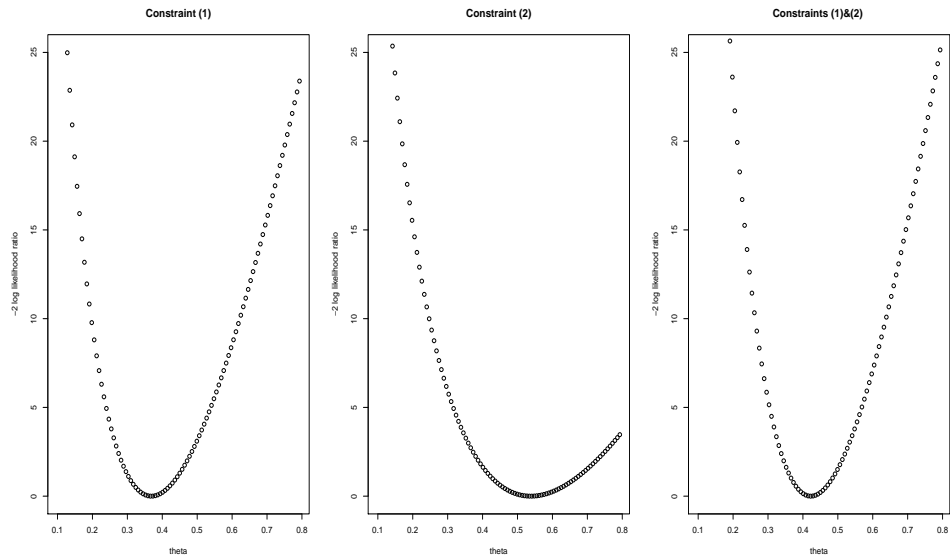


Figure 6.12: Empirical likelihood ratio statistic with different constraints and censoring distribution: 20+Exponential ($\lambda_1 = 0.02$), $n=100$

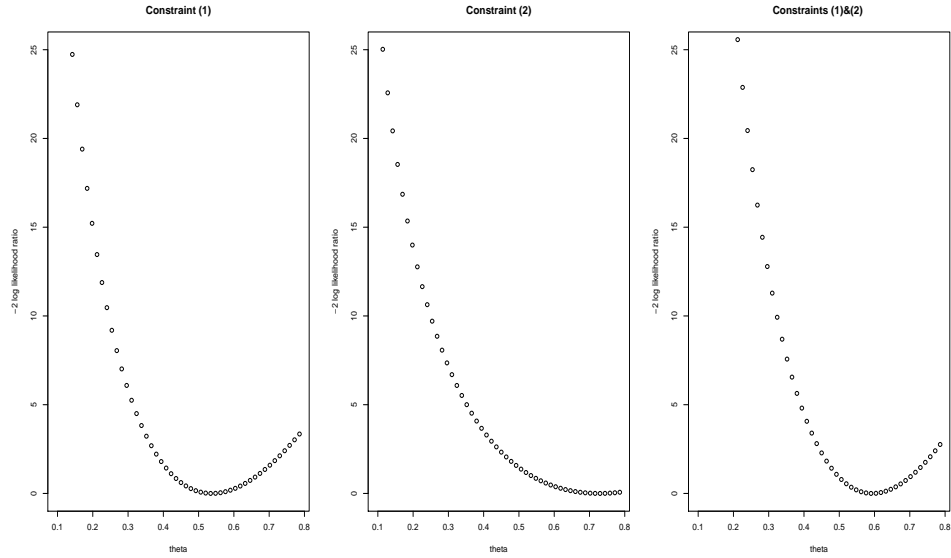


Figure 6.13: Empirical likelihood ratio statistic with different constraints and censoring distribution: 20+Exponential ($\lambda_1 = 0.02$), $n=50$

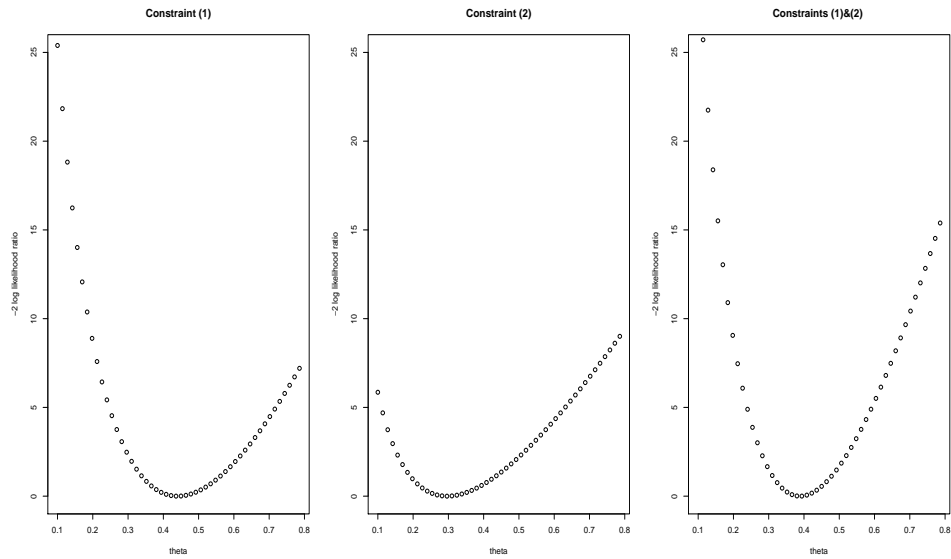


Figure 6.14: Empirical likelihood ratio statistic with different constraints and censoring distribution: 20+Exponential ($\lambda_1 = 0.02$), $n=50$

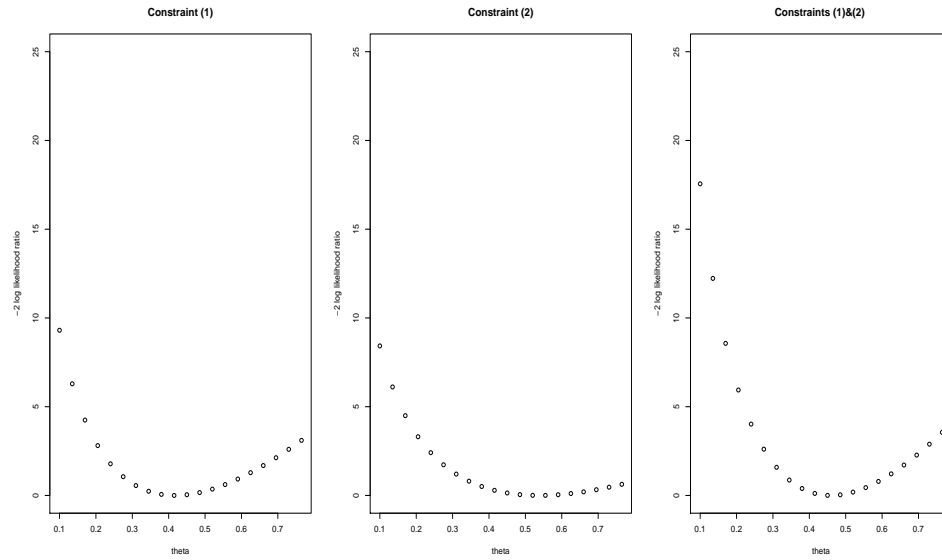


Figure 6.15: Empirical likelihood ratio statistic with different constraints and censoring distribution: 20+Exponential ($\lambda_1 = 0.02$), $n=20$

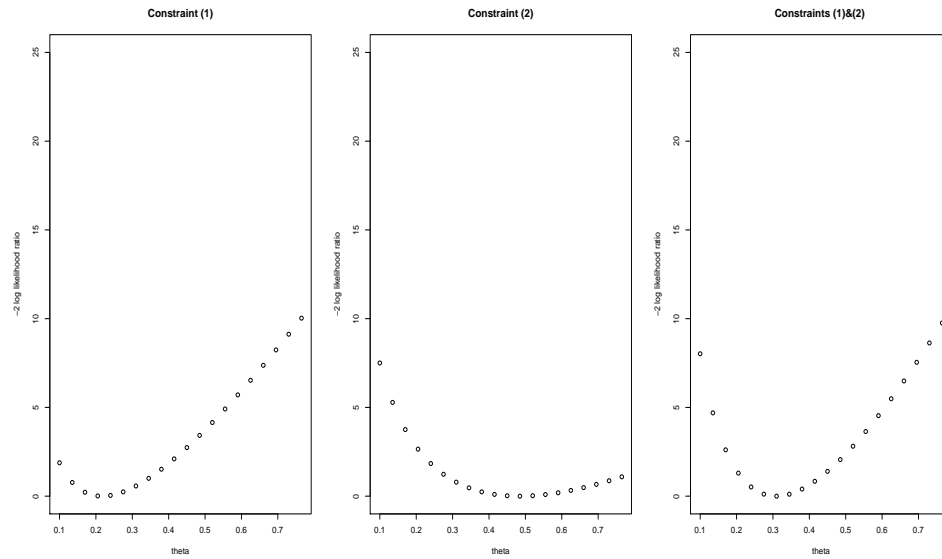


Figure 6.16: Empirical likelihood ratio statistic with different constraints and censoring distribution: 20+Exponential ($\lambda_1 = 0.02$), $n=20$

6.2 Examples

6.2.1 Example 1.

Consider the AFT model. For $i = 1, \dots, n$, let T_i be the failure time for the i th subject and let \mathbf{X}_i be the associated p -vector of covariates. The accelerated failure time model specifies that

$$\log T_i = \boldsymbol{\beta}_0^\top \mathbf{X}_i + \epsilon_i, \quad i = 1, \dots, n.$$

where $\boldsymbol{\beta}_0$ is a p -vector of unknown regression parameters and $\epsilon_i (i = 1, \dots, n)$ are independent error terms with a common, but completely unspecified, distribution.

Due to censoring, we only observe

$$\tilde{T}_i = \min(T_i, C_i), \Delta_i = I_{\{T_i \leq C_i\}},$$

where C_i are the censoring times for T_i .

We may use rank based procedures to estimate $\boldsymbol{\beta}$ and there are several available. See Chapter 7 of Kalbfleisch and Prentice (2002) [13] and Jin *et al.* (2003) [12]. Define $e_i(\mathbf{b}) = \log \tilde{T}_i - \mathbf{b}^\top \mathbf{X}_i$ and $Y_i(\mathbf{b}; t) = I_{\{e_i(\mathbf{b}) \geq t\}}$. Write $S^{(0)}(\mathbf{b}; t) = n^{-1} \sum_{i=1}^n Y_i(\mathbf{b}; t)$ and $S^{(1)}(\mathbf{b}; t) = n^{-1} \sum_{i=1}^n Y_i(\mathbf{b}; t) X_i$.

A rank based estimator of $\boldsymbol{\beta}$ can be defined as the solution to the following equations

$$\mathbf{0} = \sum_{i=1}^n \Delta_i \phi_i [\mathbf{X}_i - \bar{\mathbf{X}}\{\mathbf{b}; e_i(\mathbf{b})\}],$$

where $\bar{\mathbf{X}}(\boldsymbol{\beta}; t) = S^{(1)}(\boldsymbol{\beta}; t)/S^{(0)}(\boldsymbol{\beta}; t)$, and $\phi_i > 0$ is some weights.

We get different roots/estimators when we have different ϕ . The two choices of ϕ that Jin *et al.* (2003) used are $\phi_i \equiv 1$ and $\phi_i \equiv S^{(0)}(e_i)$, corresponding to the log-rank

and Gehan statistics, respectively. We shall also use these two as examples.

As suggested in Zhou (2005) [35], the above estimating equations will become constraint equations when maximizing the empirical likelihood function. So, we shall have the following two constraint equations corresponding to $\phi_i \equiv 1$ and $\phi_i \equiv S^{(0)}(e_i)$, respectively.

$$\mathbf{0} = \sum_{i=1}^n \Delta_i [\mathbf{X}_i - \bar{\mathbf{X}}\{\mathbf{b}; e_i(\mathbf{b})\}] \frac{w_i}{\hat{w}_i} \quad (6.1)$$

$$\mathbf{0} = \sum_{i=1}^n \Delta_i S^{(0)}(e_i) [\mathbf{X}_i - \bar{\mathbf{X}}\{\mathbf{b}; e_i(\mathbf{b})\}] \frac{w_i}{\hat{w}_i} \quad (6.2)$$

where the w_i are the jumps of hazard function that also appear in the definition of the log AL :

$$\log AL(\mathbf{w}, b) = \sum_{i=1}^n \delta_i \log w_i - \sum_{i=1}^n w_i R_i .$$

and \hat{w}_i are the jump sizes of the Nelson-Aalen estimator computed from $\delta_i, r_i(b)$. Apparently (6.1) and (6.2) are functionally independent.

If we only use constraint equations (6.1) or (6.2), it is called the just determined case, because the number of estimating/constraint equations equals the number of unknown parameters. In many cases, however, it is desirable to use the two system of estimating equations (6.1) and (6.2) simultaneously so this becomes an over-determined system of equations: p parameters (= dim of $\boldsymbol{\beta}$), $2p$ equations.

The data we are going to use is the Myeloma data, which was used by Jin *et al.* (2003) and Zhou (2005). The data set includes 65 subjects, with 17 of them having right censored survival times. The data is available inside the latest version of the package **emplik** in **R**.

We work with the log of the survival times, and fitting the model

$$\log(\text{time}) = \beta \log(\text{BUN}) + \epsilon$$

with both constraint equations (6.1) and (6.2) together as over-determined constraint equations.

The maximum empirical likelihood estimator $\hat{\beta} = -1.524$; with a 95% confidence interval $[-2.49, -0.65]$. If we use just one constraint equations of Gehan (the just-determined case), we get an estimator of -1.685 , and a 95% confidence interval is $[-2.67, -0.58]$. If we use just one constraint equations of log-rank, we get an estimator of -1.678 , and a 95% confidence interval is $[-2.84, -0.22]$.

We see that the maximum empirical likelihood estimators are close to each other, but confidence interval is the shortest if we use both equations in the over-determined case, which is also proved by our Corollary.

Jin *et al.* (2003) also compared the Gehan estimator and the log-rank estimator in different scenarios, and obviously the two estimators have their own advantages (smaller bias or smaller standard deviation) in different situations. The performance of the estimator depends on the distribution of ϵ which is unknown in practice, so they suggested to adopt the approach of Lai & Ying by constructing data-dependent weight function ϕ .

Using our over-determined result, we do not have to choose between the estimators, we can obtain an estimator by using both (or even more) of those constraint equations, as long as they are functionally independent. We can get an estimator by using our theory, which says this estimator is better than either one asymptotically (having smaller variance asymptotically) and a χ^2 based likelihood ratio test is avail-

able.

6.2.2 Example 2.

This is the same example as the one in Kim (2003) [14] dissertation.

An AML study by Embury *et al.* at Stanford University reports the results of a clinical trial to evaluate the efficacy of maintenance chemotherapy for acute myelogenous leukemia (AML). There are two groups: one with maintenance chemotherapy and the other without.

Based on the Kaplan-Meier survival curves drawn from the data, we are convinced that a hybrid model is appropriate to fit the data:

$$1 - G(t) = [1 - F(t - \theta)]^\eta, \quad \text{for any } t \in \mathfrak{R}^1.$$

where $G(t)$ and $F(t)$ are two unknown CDF for survival times from the two different groups.

This is a special case of our research. There are only two parameters θ and η , but $r > 2$ constraints since we assume the data fit the hybrid model.

$$\sum_j \delta_{y_j} g_k(y_j) \log(1 - v_j) = \sum_i \delta_{x_i} \eta g_k(x_i - \theta) \log(1 - w_i), \quad k = 1, \dots, r.$$

where $g_k, k = 1, \dots, r$ are functionally independent functions satisfy some conditions, (y_j, δ_{y_j}) and (x_i, δ_{x_i}) are censored observations from two samples, and v_j and w_i are hazard jumps from the corresponding two samples.

Chapter 7 Discussions and Future Research Questions

Empirical likelihood in terms of distribution function with uncensored-data has been widely studied by many people. Owen's 2001 book contains many important results. However, for censored data, there are still many problems to investigate.

In Chapter 3, we extend Qin and Lawless's results to right-censored data for the case where the number of estimating equations is larger than the number of parameters. In Chapter 4, we present a mathematical derivation of empirical likelihood for two-sample mixed hypothesis similar to Owen, but for right-censored data and using hazard function. Hazard-type empirical likelihood and constraints are much easier to handle for censored-data than distribution-type constraints, so we try to establish a general relationship between the constrained Kaplan-Meier estimator and the corresponding constrained Nelson-Aalen estimator in Chapter 5.

The results presented in this dissertation are primarily theoretical. Computations associated with some of these results are challenging and need further investigations. We focus on right-censored data. A general result for left-censored data, interval-censored data and truncation data might be further studied.

As Qin and Lawless remarked, a good deal of work is needed to apply and access the above methods in practical situations. Experience is needed to determine how easily estimates can be obtained in small- to moderate-size samples and what the properties of the estimators and the empirical-likelihood-ratio statistics are in these situations. These topics can be further investigated.

Appendix: Program Code

Annotated code for over-determined log-likelihood

```
newdataclean<-function (x, d, y = -Inf, fun, itertrace = FALSE) {
#Store data x as a vector.
  x <- as.vector(x)
#Store length of x.
  n<-length(x)
#Check that there are at least 3 data in x.
  if (n <= 2)
    stop("Need more observations in x")
#Check that status and observations have same length in x.
  if (length(d) != n)
    stop("length of x and d must agree")
#Check that status are only 0, 1.
  if (any((d != 0) & (d != 1)))
    stop("d must be 0/1's for censor/not-censor")
#Check that x are numeric (for example, no NA values).
  if (!is.numeric(x))
    stop("x must be numeric -- observed times")
#"Clean" data using function Wdataclean2.
  newdata <- Wdataclean2(z=x, d = d)
#Further "clean" data using function DnR.
  temp <- DnR(newdata$value, newdata$dd, newdata$weight,y = y)
#Calculate the Nelson-Aalen jumps.
```

```

    jump <- (temp$n.event)/temp$n.risk
#Calculate the constraint functions by plugging in the observations.
    funtime <- as.matrix(fun (temp$times))
#Check the dimension of the output of the constraint functions.
    if (ncol(funtime) !=2)
        stop("check the output dim of fun ")
#Calculate  $\hat{\theta}$  by using Nelson-Aalen jumps.
    esttheta <- t(jump) %*% funtime
    if (itertrace)
        print(c("thetahat=", esttheta))
    ltheta<-min(esttheta[1], esttheta[2])
    rtheta<-max(esttheta[1], esttheta[2])
#Store the jumps that are less than 1.
    index<- (jump < 1)
    K12<-rep(0,2)
#Store the observations that have jumps equal to 1.
    tm1<- temp$times[!index]
#Check the length of tm1, which should not be greater than 1.
    if (length(tm1) > 1)
        stop("more than 1 place jump>=1 in x?")
#If the last jump is 1, recalculate K12.
    if (length(tm1) > 0) {
        K12 <- K12 + as.vector(fun (tm1))}
#Restore the following variables with jumps less than 1.
    eve <-temp$n.event[index]
    tm <- temp$times[index]
    rsk<- temp$n.risk[index]

```

```

    jmp <- jump [index]
    funt <- as.matrix(fun (tm))
    list(funt=funt, eve=eve, rsk=rsk, ltheta=ltheta, rtheta=rtheta,
        K12=K12,n=n, jump=jump, tm=tm)
}

```

#This function is to calculate the constraints.

```

newgradf<-function(lam, funt, eve, rsk, K, n) {
  arg<- as.vector(rsk + funt%% lam)
  VV <- (eve * llogp(arg, 1/n)) %% funt - K
  return(as.vector(VV))
}

```

```

newloglik<-function(funt, eve, rsk, n, maxit = 25, K12, theta, tola
= 1e-07, itertrace = FALSE){

```

```

  TINY <- sqrt(.Machine$double.xmin)
  if (tola < TINY)
    tola <- TINY
  lam <- rep(0,2)

```

#Newton-Raphson process.

```

  nwts <- c(3^-c(0:3), rep(0, 12))
  gwts <- 2^(-c(0:(length(nwts) - 1)))
  gwts <- (gwts^2 - nwts^2)^0.5
  gwts[12:16] <- gwts[12:16] * 10^-c(1:5)
  nits <- 0
  gsize <- tola + 1

```

```

while (nits < maxit && gsize > tola) {
  grad <- newgradf(lam, funt, eve, rsk, K = theta - K12, n = n)
  gsize <- mean(abs(grad))
  arg <- as.vector(rsk + funt %*% lam)
  ww <- as.vector(-llogpp(arg, 1/n))^0.5
  tt <- sqrt(eve) * ww
  HESS <- - (t(funt * tt) %*% (funt * tt) )
  nstep <- as.vector(-solve(HESS, grad))
  gstep <- grad
  if (sum(nstep^2) < sum(gstep^2))
    gstep <- gstep * (sum(nstep^2)^0.5/sum(gstep^2)^0.5)
  ninner <- 0
  for (i in 1:length(nwts)) {
    lamtemp <- lam + nwts[i] * nstep + gwts[i] * gstep
    ngrad <- newgradf(lamtemp, funt, eve, rsk, K = theta - K12, n = n)
    ngsiz <- mean(abs(ngrad))
    if (ngsiz < gsiz) {
      lam <- lamtemp
      ninner <- i
      break
    }
  }
  nits <- nits + 1
  if (ninner == 0)
    nits <- maxit
  if (itertrace)
    print(c(lam, gsiz, ninner))
}

```



```

    }
#Calculate the log-likelihood.
    lamfun<-as.vector(funt%%lam)
    onePlamf<-(rsk+lamfun)/rsk
    nloglik<- sum(eve*llog(onePlamf*rsk,1/n))+sum(eve*llogp(onePlamf, 1/n))
    list(nloglik =nloglik, lambda =lam) }

```

Code for Simulation 1

```

estfun<-function(x){
  cbind(as.numeric(x<=20),as.numeric(20<=x&x<=40)) }
llikratio<-rep(NA,1000)
dloglik<-rep(NA,1000)
nuloglik<-rep(NA,1000)
for (j in 1:1000){
  t<-rexp(50,0.02)
  d<-rexp(50,0.005)
  x<-pmin(t,d)
  cen<-as.numeric(t<=d)
  newtemp<-newdataclean(x=x,d=cen, fun=estfun)
  nuloglik[j]<-newloglik(newtemp$funt, newtemp$eve, newtemp$rsk,
  newtemp$n, maxit = 25, newtemp$K12,0.4)$nuloglik
  testthetas<-matrix(NA, ncol=101, nrow=1000)
  testloglik<-matrix(NA,ncol=101, nrow=1000)
  for (i in 1:101){
    testthetas[j,i]<-newtemp$ltheta+(i-1)/100*(newtemp$rtheta-newtemp$ltheta)
    testloglik[j,i]<- newloglik(newtemp$funt, newtemp$eve, newtemp$rsk,

```

```

newtemp$n, maxit = 25, newtemp$K12, testthetas[j,i])$nloglik }
dloglik[j]<- max(-testloglik[j,])

llikratio[j]<- 2*(dloglik[j]+nuloglik[j]) }
plot(qchisq(1:1000/1000,1), sort(llikratio), xlab= "chisq(1)
quantiles ", ylab= "-2 log likelihood ratio ") abline(a=0,b=1)

```

Code for Simulation 2

```

t<-rexp(100,0.02)
d<-rexp(100,0.005)
x<-pmin(t,d)
cen<-as.numeric(t<=d)
thetas1<-rep(NA,100)
result1<-rep(NA,100)
est1<- function(x){ as.numeric(x<=20)}
result2<-rep(NA,100)

est2<- function(x){as.numeric(20<=x & x<=40)} for (i in 1:100){
thetas1[i]<- 0.1+(i-1)*0.7/100
result1[i]<-empH1.test(x=x,d=cen,y=-Inf,theta=thetas1[i],fun=est1,
tola=.Machine$double.eps^.25)$`-2LLR`
result2[i]<-empH1.test(x=x,d=cen,y=-Inf,theta=thetas1[i],fun=est2,
tola=.Machine$double.eps^.25)$`-2LLR`}
thetas2<-rep(NA,100)
result<-rep(NA,100)
thetas2<-rep(NA,100)

```

```

nuloglik<-rep(NA,100) dloglik<-rep(NA,100)
testthetas<-matrix(NA,ncol=101,nrow=100)
testloglik<-matrix(NA,ncol=101,nrow=100)
newtemp<-newdataclean(x=x,d=cen, fun=estfun)
for (i in 1:100){
thetas2[i]<- 0.1+(i-1)*0.7/100

nuloglik[i]<- newloglik(newtemp$fun,newtemp$eve, newtemp$rsk,
newtemp$n, maxit = 25, newtemp$K12,thetas2[i])$nuloglik

for (j in 1:101){
testthetas[i,j]<-newtemp$ltheta+(j-1)/100*(newtemp$rtheta-newtemp$ltheta)
testloglik[i,j]<- newloglik(newtemp$fun, newtemp$eve, newtemp$rsk,
newtemp$n, maxit = 25, newtemp$K12, testthetas[i,j])$nloglik }
dloglik[i]<- max(-testloglik[i,])

result[i]<- 2*(dloglik[i]+nuloglik[i])}
par( mfrow = c(1,3))
plot(thetas1,result1,ylim=c(0,25),xlab="theta",ylab="-2 log
likelihood ratio",main="Constraint (1)")
plot(thetas1,result2,ylim=c(0,25),xlab="theta",ylab="-2 log
likelihood ratio",main="Constraint (2)")
plot(thetas2,result,ylim=c(0,25),xlab="theta",ylab="-2 log
likelihood ratio", main="Constraints (1)&(2) ")

```

Copyright© Yanling Hu, 2011.

Bibliography

- [1] Akritas, M.G. (2000). "The Central Limit Theorem Under Censoring." *Bernoulli*, **6**, pp. 1109-1120.
- [2] Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer Verlag, New York.
- [3] Casella, G. and Berger, R. L. (2002). *Statistical Inference*, Duxbury Press, Pacific Grove, California.
- [4] Chen M. (2005). "Some Contributions To The Empirical Likelihood Method." Ph.D Dissertation, Department of Statistics, University of Kentucky.
- [5] DiCiccio, T.J., Hall, P. and Romana, J.P. (1989). "Comparisons Of Parametric And Empirical Likelihood Functions." *Biometrika*, **76**, pp. 465-476.
- [6] DiCiccio, T.J., Hall, P. and Romana, J.P. (1991). "Empirical Likelihood Is Bartlett-Correctable." *The Annals of Statistics*, **19**, pp. 1053-1061.
- [7] DiCiccio, T.J. and Romana, J.P. (1989). "On Adjustments To The Signed Root Of The Empirical Likelihood Ratio Statistic." *Biometrika*, **76**, pp. 447-456.
- [8] Fang, H. (2000). "Binomial Empirical Likelihood Ratio Method In Survival Analysis." Ph.D Dissertation, Department of Statistics, University of Kentucky.
- [9] Graybill, F. (1976). *Theory And Application Of The Linear Model*, Wadsworth Publishing Company Inc., Belmont, California.

- [10] Hall, P. (1990). "Pseudo-likelihood Theory For Empirical Likelihood." *Ann. Statist.*, **18**, pp. 121-140.
- [11] Hoeffding, W. (1948). "A Class of Statistics with Asymptotically Normal Distribution." *Ann. Math. Statist.*, **19**, pp. 293-325.
- [12] Jin, Z., Lin, D.Y., Wei, L.J. and Ying, Z. (2003). "Rank-based Inference For The Accelerated Failure Time Model." *Biometrika*, **90**, pp. 341-353.
- [13] Kalbfleisch, J.D. and Prentiss, R.L. (2002). *The Statistical Analysis of Failure Time Data*. Wiley-Interscience, Hoboken, New Jersey.
- [14] Kim, K. (2003). "Empirical Likelihood Ratio Method When Additional Information Is Known." Ph.D Dissertation, Department of Statistics, University of Kentucky.
- [15] Kim, K. and Zhou, M. (2004). "Symmetric Location Estimation/Testing By Empirical Likelihood." *Communications in Statistics: Theory and Methods*, **33**, pp. 2233-2243.
- [16] Kolaczyk, E.D. (1992). "Empirical Likelihood For Generalized Linear Models." Technical Report 389, Department of Statistics, Stanford University.
- [17] Li, G. (1995). "On Nonparametric Likelihood Ratio Estimation Of Survival Probabilities For Censored Data." *Statist. and Prob. Letters*, **25**, pp. 95-104.
- [18] Li, G., Li, R. and Zhou, M. (2005). "Empirical Likelihood In Survival Analysis." In *Contemporary Multivariate Analysis and Design of Experiments*, edited by J. Fan and G. Li., World Scientific Publisher, New Jersey, pp. 337-350.
- [19] Liu, L. (2008). private notes.

- [20] Murphy, S.A. (1995). "Likelihood Ratio-Based Confidence Intervals In Survival Analysis." *J. Amer. Stat. Assoc.*, **90**, pp. 1399-1405.
- [21] Murphy, S.A. and Van der Vaart, A. W. (1997). "Semiparametric Likelihood Ratio Inference." *Ann. Statist.*, **25**, pp. 1471-1509.
- [22] Owen, A.B. (1988). "Empirical Likelihood Ratio Confidence Intervals For A Single Functional." *Biometrika*, **75**, pp. 237-249.
- [23] Owen, A.B. (1990). "Empirical Likelihood Ratio Confidence Regions." *Ann. Stat.*, **18**, pp. 90-120.
- [24] Owen, A.B. (1991). "Empirical Likelihood For Linear Models." *Ann. Stat.*, **19**, pp. 1725-1747.
- [25] Owen, A.B. (1992). "Empirical Likelihood And Generalized Projection Pursuit." Technical Report 393, Department of Statistics, Stanford University.
- [26] Owen, A.B. (2001). *Empirical Likelihood*. Chapman & Hall/CRC, Boca Raton.
- [27] Pan, X. R. (1997). "Empirical Likelihood Ratio Method For Censored Data." Ph.D Dissertation, Department of Statistics, University of Kentucky.
- [28] 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.
- [29] Pan, X. R. and Zhou, M. (2002). "Empirical Likelihood Ratio In Terms Of Cumulative Hazard Function For Censored Data." *J. Multivar. Anal.*, **80**, pp. 166-188.
- [30] Qin, J. and Lawless, J. (1994). "Empirical Likelihood And General Estimating Equations." *Ann. Stat.*, **22**, pp. 300-325.

- [31] 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.
- [32] Wilks, S. S. (1938). “The Large-sample Distribution Of The Likelihood Ratio For Testing Composite Hypotheses.” *Ann. Math. Statist.*, **9**, pp. 60-62.
- [33] Zhou, M. (1991). “Some Properties of the Kaplan-Meier Estimator for Independent Nonidentically Distributed Random Variables.” *Ann. Stat.*, **19**, pp. 2266-2274.
- [34] Zhou, M. (2000). “Empirical Envelope MLE And LR Test.” Technical Report, Department of Statistics, University of Kentucky.
- [35] Zhou, M. (2005). *private notes*.

Vita

Name

- Yanling Hu

Date of Birth

- 1982

Place of Birth

- Qingdao, Shandong, China.

Education

- 2001-2005, B.S. Statistics, Fudan University, Shanghai, China

Employment

- 2006-2008, Teaching Assistant, University of Kentucky, Lexington, KY
- 2008-2011, Research Assistant, University of Kentucky, Lexington, KY

Honors

- R.L. Anderson Award for Outstanding Teaching, Department of Statistics, University of Kentucky (2008)
- ENAR Travel Award, Fostering Diversity in Biostatistics Workshop

(2010)

- R.L. Anderson Award for Best Research, Department of Statistics, University of Kentucky (2010)
- Finalist of The Student Paper Competition for The ASA Nonparametric Statistics Section, JSM (2010)

Copyright© Yanling Hu, 2011.