2011

Estimation For The Cox Model With Various Types Of Censored Data

Tonya Riddlesworth
University of Central Florida

Estimation for the Cox Model with Various Types of Censored Data

by

Tonya Riddlesworth
M.S. Mathematics, University of Central Florida, 2006

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Mathematics
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Fall Term
2011

Major Professor: Jian-Jian Ren

ABSTRACT

In survival analysis, the Cox model is one of the most widely used tools. However, up to now there has not been any published work on the Cox model with complicated types of censored data, such as doubly censored data, partly-interval censored data, etc., while these types of censored data have been encountered in important medical studies, such as cancer, heart disease, diabetes, etc. In this dissertation, we first derive the bivariate nonparametric maximum likelihood estimator (BNPMLE) $\hat{F}_n(t,z)$ for joint distribution function $F_0(t,z)$ of survival time T and covariate Z, where T is subject to right censoring, noting that such BNPMLE $\hat{F}_n$ has not been studied in statistical literature. Then, based on this BNPMLE $\hat{F}_n$ we derive empirical likelihood-based (Owen, 1988) confidence interval for the conditional survival probabilities, which is an important and difficult problem in statistical analysis, and also has not been studied in literature. Finally, with this BNPMLE $\hat{F}_n$ as a starting point, we extend the weighted empirical likelihood method (Ren, 2001 and 2008a) to the multivariate case, and obtain a weighted empirical likelihood-based estimation method for the Cox model. Such estimation method is given in a unified form, and is applicable to various types of censored data aforementioned.

Key Words and Phrases: Bivariate Maximum Likelihood Distribution Estimator, Conditional Survival Probability, Cox Model, Doubly Censored Data, Interval Censored Data, Maximum Likelihood Estimator, Partly Interval-Censored Data, Right Censored Data, Weighted Empirical Likelihood

To my wonderful mom, Christine

ACKNOWLEDGMENTS

I wish to extend my deepest appreciation to my Ph.D. advisor, Professor Jian-Jian Ren, for her excellent guidance and invaluable advice she has given me throughout the years. The past few years have been full of struggle and frustration, but she trusted in me and encouraged me to keep working and for that I will be forever grateful. I also want to thank my dissertation committee members, Professors Ram Mohapatra, Liqiang Ni, Gary D. Richardson and James R. Schott, for all their advice, support and time spent devoted to my studies at UCF. I would like to send a special thank you to Professor Mohapatra and the chair of the Math Department, Professor Mikusinski, who have helped me tremendously through some difficult situations.

In addition, this dissertation would never have been completed without the devotion of my family and friends, particularly my dear fiance, Patrick Gant, who has been a great friend to lean on and has shown amazing patience, encouragement and understanding. Finally, words alone cannot express the gratitude, admiration and love I have for my mom, Christine Riddlesworth. I dedicate this dissertation to her.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1. INTRODUCTION AND PRELIMINARIES

The Cox model is one of the most widely used tools in analysis of survival data. But up to now there has not been any published work on the Cox model with complicated types of censored data, such as doubly censored data, partly interval-censored data, etc. In this dissertation, weighted empirical likelihood (Ren, 2001) is used to develop general estimation methods for the Cox model with the various types of censored data aforementioned. To facilitate this work, we derive the bivariate maximum likelihood distribution estimator for right censored data, which also leads to the construction of the empirical likelihood-based confidence interval for conditional survival probabilities.

1.1 Introduction

Survival analysis is a branch of statistics concerned with the *failure time*, or the event time, i.e., the time elapsed from a specific time origin until a failure occurs. In practice, a failure time could be, for example, the age when a child learns a certain task, the time of death from a particular disease, etc. Often, the interest of study is to determine the effect of certain independent variables, called *covariates*, on the failure time of a subject. In particular, medical researchers are often interested in the effect of a covariate, such as a treatment, on a patient's survival time. Sometimes traditional methods, such as multiple regression, can be used to determine the effect of the covariates on the failure time. However, two common situations arise where these methods cannot be used. First, when the failure time is not normally distributed, conventional methods such as least squares multiple regression cannot be used. Second, traditional methods break down when censored data are observed. In this context, the Cox model (Cox, 1972) provides a useful tool because it does not require the underlying distribution to be normal and can include right censored data in its estimation procedures.

So far, most methods developed for the Cox model deal with right censored data. But in recent years, more complicated types of censored data, such as doubly censored data, interval censored data, partly interval-censored data, etc., have been encountered in important scientific studies, including cancer research, AIDS research, heart and diabetes disease research, etc. There have been some papers that extend the Cox model to deal with interval censored data (Satten, 1996; Huang, 1996; among others), but for other types of censored data mentioned above the statistical methods on the Cox model are lacking. This work will develop general estimation methods for the Cox model with various types of censored data by utilizing the weighted empirical likelihood method (Ren, 2001).

Likelihood is the most widely used procedure for inference in parametric as well as nonparametric models. One reason for this is that the estimators usually possess desirable asymptotic properties. The usual *maximum likelihood estimator* is obtained by maximizing the *parametric likelihood function* and is shown to be usually consistent, efficient, and asymptotically normal (Casella and Berger, 2002). Wilks (1938) showed that, under some conditions, the logarithm of the parametric likelihood ratio has an asymptotic chi-square distribution, which provides the key to constructing confidence sets for the parameter. To overcome the limitations of parametric likelihood, Owen (1988) proposed *empirical likelihood*, which is a nonparametric method. Over the past two decades, it has been shown (see Owen, 1988, 1990, 1991; DiCiccio *et al.*, 1991; Qin and Lawless, 1994; Mykland, 1995; among others) that Wilks' theorem holds for empirical likelihood and that empirical likelihood inferences are of comparable accuracy to alternative methods. In particular, DiCiccio, Hall and Romana (1991) showed that empirical likelihood is Bartlett-correctable. However, when imposing a model assumption with those complicated types of censored data above mentioned the empirical likelihood function is usually very complicated and very difficult or impossible to maximize. To deal with these issues, Ren (2001, 2008a) proposed *weighted empirical likelihood*, which provides a simple and direct way to incorporate some model assumptions in the derivation of the likelihood function for various types of censored data. So far, the results on weighted empirical likelihood have shown to be favorable to alternative methods, but weighted empirical likelihood has only been applied to the univariate case.

In this dissertation, we extend the weighted empirical likelihood method to multivariate survival data, and we develop a general weighted empirical likelihood-based estimation method for the Cox model with various types of censored data aforementioned. Specifically, we derive the weighted empirical likelihood-based estimating equation(s) for the regression parameters in the Cox model. To facilitate this, we first derive the bivariate maximum likelihood distribution estimator for right censored data. In addition, we derive the empirical likelihood-based confidence interval for conditional survival probabilities with right censored data.

This chapter is organized as follows: Section 1.2 gives a brief review of the *proportional hazards model* and, in particular, the Cox model; Section 1.3 introduces several types of censored data, gives real data examples for each type of censored data, and briefly reviews some relevant asymptotic results; Section 1.3 also reviews some recent results on the Cox model; Section 1.4 reviews statistical inference based on parametric likelihood, empirical likelihood, and weighted empirical likelihood; and Section 1.5 summarizes the main results of this dissertation.

1.2 Cox's Proportional Hazards Model

Cox's proportional hazards model (in short, the Cox model) is a commonly used tool in survival analysis to explore the relationship between the failure time and covariates. The Cox model is a specific form of the proportional hazards model, and is popular because it assumes that the risk for the treatment is proportional to the risk for the control, but makes no assumptions on the shape of the underlying distribution function. In some limited situations, the assumption of proportional hazards (or risks) can be tested via goodness-of-fit tests, while in some fields there is empirical evidence to support the idea of proportional hazards. Another benefit of the Cox model is that the model is sufficiently flexible to incorporate covariates and deal with right censored data (see description of right censored data in Section 1.3.1). Next, we review some basic definitions and concepts in survival analysis, then give the definitions of the proportional hazards model and the Cox model in Sections 1.2.1-1.2.2, respectively.

Throughout this dissertation, let T be a non-negative continuous random variable, the failure time variable, with distribution function (d.f.) $F_T(t)$ and density function $f_T(t)$. The *hazard function* for T is the instantaneous risk, i.e., the probability that the failure occurs at time $T = t$ given survival up to time t:

$$h_T(t) = \lim_{\Delta \to 0+} \frac{P\{t < T \leq t + \Delta \,|\, T > t\}}{\Delta}. \tag{1.2.1}$$

The hazard function is especially valuable in survival analysis because it provides a way to consider the immediate risk attached to a subject.

In order to obtain an expression for the density function $f_T(t)$ in terms of the hazard function $h_T(t)$, we utilize the definitions of derivatives and conditional probability to obtain:

$$\begin{aligned} h_T(t) = \lim_{\Delta \to 0+} \frac{P\{t < T \leq t + \Delta \,|\, T > t\}}{\Delta} \;&=\; \lim_{\Delta \to 0+} \frac{P\{t < T \leq t + \Delta\}}{\Delta \cdot P\{T > t\}} \\ = \lim_{\Delta \to 0+} \frac{F_T(t+\Delta) - F_T(t)}{\Delta \cdot \bar{F}_T(t)} \;&=\; \frac{F_T'(t)}{\bar{F}_T(t)} \;=\; \frac{f_T(t)}{\bar{F}_T(t)}, \end{aligned} \tag{1.2.2}$$

where $\bar{F}_T(t) = 1 - F_T(t) = P\{T > t\}$ is the *survivor function* for T. Note that from $f_T(t) = -\bar{F}_T'(t)$, we have

$$h_T(t) = -\frac{d}{dt}\Big(\log \bar{F}_T(t)\Big), \tag{1.2.3}$$

and since $\bar{F}_T(0) = 1$, we integrate (1.2.3) and obtain:

$$\bar{F}_T(t) = \exp\left\{-\int_0^t h_T(u)du\right\} = \exp\{-H_T(t)\}, \tag{1.2.4}$$

where $H_T(t) = \int_0^t h_T(u)du$ is the *cumulative hazard function* of T. Taking derivatives in (1.2.4), we have

$$f_T(t) = h_T(t)\exp\{-H_T(t)\}. \tag{1.2.5}$$

Equations (1.2.4)-(1.2.5) give the relationship among $f_T(t)$, $\bar{F}_T(t)$, and $h_T(t)$. The hazard function, $h_T(t)$, is used in the following subsections to define the proportional hazards model and the Cox model.

1.2.1 Proportional Hazards Model

As follows, we describe the proportional hazards (PH) model, by first reviewing the two-sample problem, then extending it to the general form of the PH model.

Two-Sample Problem

Consider two samples: a control group and a treatment group. Let $h_{T0}(t)$ and $h_{T1}(t)$ be the hazard functions of the control group and the treatment group, respectively, and denote the two observed samples as follows:

$$\begin{cases} \text{Control Group :} & T_1^{(0)}, \cdots, T_{n_0}^{(0)} \\ \text{Treatment Group :} & T_1^{(1)}, \cdots, T_{n_1}^{(1)}. \end{cases} \tag{1.2.6}$$

The PH model assumes that at time t, the hazard function $h_{T0}(t)$ of the control group and the hazard function $h_{T1}(t)$ of the treatment group are proportional up to a constant $\psi > 0$ and is written as:

$$h_{T1}(t) = \psi h_{T0}(t). \tag{1.2.7}$$

Note that the value of ψ determines whether the treatment is effective (i.e., better than the control). If $\psi < 1$ in (1.2.7), then at any given time t, the hazard (or risk) for the treatment is less than that for the control, thus the treatment is effective. On the other hand, if $\psi > 1$ in (1.2.7), then at any given time t, the hazard for the treatment is greater than that for the control, thus the treatment is not effective. Some procedures are available (which will be described in Section 1.2.2) to check whether the treatment is effective by testing

$$H_0 : \psi = 1 \quad \text{vs.} \quad H_1 : \psi < 1. \tag{1.2.8}$$

For two-sample case (1.2.7)-(1.2.8), if we introduce a dichotomous variable Z, with

$$Z = \begin{cases} 0 & \text{control group} \\ 1 & \text{treatment group,} \end{cases} \tag{1.2.9}$$

and let $h(t; z)$ denote the conditional hazard function of the survival time T given $Z = z$, then we have $h_{T1}(t) = h(t; 1)$ and $h_{T0}(t) = h(t; 0)$, in turn, PH model (1.2.7) becomes

$$h(t; z) = \psi(z) h_{T0}(t), \quad \text{where} \;\; \psi(z) = \begin{cases} 1 & z = 0 \\ \psi & z = 1. \end{cases} \tag{1.2.10}$$

The observed data in this case is written as

$$(T_1, Z_1), \cdots, (T_n, Z_n), \tag{1.2.11}$$

where $n = n_0 + n_1$, $T_i = T_j^{(0)}$ or $T_k^{(1)}$, and $Z_i = 0$ or 1 for $i = 1, \cdots, n$.

Generalization of Two-Sample PH Model

PH model (1.2.10) can be generalized for cases when Z has multiple possible values $z_0, z_1, , \cdots, z_k$. For example, suppose z_i represents a different treatment for a particular disease, i.e.,

$$Z = \begin{cases} 0 & \text{control} \\ 1 & \text{treatment 1} \\ \vdots & \qquad \vdots \\ k & \text{treatment } k. \end{cases} \tag{1.2.12}$$

Then, $h(t;j)$ is the conditional hazard function of the survival time T under the jth treatment, and given $Z=j$, PH model (1.2.10) becomes

$$h(t;j) = \psi(j) h_{T0}(t), \qquad j = 0, 1, \cdots, k, \tag{1.2.13}$$

where $\psi(j) > 0$ is a constant corresponding to the jth treatment group. Here, the observed samples are

$$\begin{cases} \text{Control Group :} & T_1^{(0)}, \cdots, T_{n_0}^{(0)} \\ \text{Treatment Group 1 :} & T_1^{(1)}, \cdots, T_{n_1}^{(1)} \\ \text{Treatment Group 2 :} & T_1^{(2)}, \cdots, T_{n_2}^{(2)} \\ \qquad \vdots & \qquad \vdots \\ \text{Treatment Group } k : & T_1^{(k)}, \cdots, T_{n_k}^{(k)}. \end{cases} \tag{1.2.14}$$

which can be written as

$$(T_1, Z_1), \cdots, (T_n, Z_n), \tag{1.2.15}$$

where $n = n_0 + n_1 + \cdots + n_k$, and for each $i = 1, \cdots, n$, we have $T_i = T_{k_j}^{(j)}$ and $Z_i = j$ for some $j = 0, 1, \cdots, k$.

Therefore, when covariate variable Z is discrete or continuous, the generalized PH model formula is given by

$$h(t;z) = \psi(z) h_0(t), \tag{1.2.16}$$

where $\psi(z) > 0$ is a function satisfying $\psi(0) = 1$, T is the survival time of the population under consideration, $h(t; z)$ is the conditional hazard function of T given $Z = z$, and $h_0(t) = h_{T0}(t)$ is the hazard function for $T^{(0)}$, which is the survival time T when $Z = 0$. In practice, the observed data for PH model (1.2.16) are

$$(T_1, Z_1), \cdots, (T_n, Z_n), \tag{1.2.17}$$

where n is the sample size, and for each $i = 1, \cdots, n$, T_i is the survival time and Z_i is the observed variable on Z.

Further Generalization of PH Model

PH model (1.2.16) can be generalized further for cases where there are multiple explanatory variables, say $\mathcal{Z}_1, \cdots, \mathcal{Z}_k$, each of which may affect the survival time T. For example, $\mathcal{Z}_1$ could be the treatment given to the patient, $\mathcal{Z}_2$ the patient's gender, $\mathcal{Z}_3$ the patients's age, etc. If we let $\boldsymbol{Z} = (\mathcal{Z}_1, \cdots, \mathcal{Z}_k)^\top$, then PH model (1.2.16) becomes

$$h(t; \boldsymbol{z}) = \psi(\boldsymbol{z}) h_0(t), \tag{1.2.18}$$

where $\psi(\boldsymbol{z}) > 0$ and $\psi(\boldsymbol{0}) = 1$, and the rest are the same as in (1.2.16). Here, the data observed in practice may be written as

$$(T_1, \boldsymbol{Z}_1), \cdots, (T_n, \boldsymbol{Z}_n), \tag{1.2.19}$$

where n is the sample size, and for the ith individual, T_i is the survival time and $\boldsymbol{Z}_i$ is the vector of explanatory variables.

In practice, if the survival time T is subject to right censoring (see more on this in Section 1.3.1), then the data in (1.2.19) is written as

$$(V_i, \delta_i, \boldsymbol{Z}_i), \qquad i = 1, \cdots, n \tag{1.2.20}$$

where $V_i = \min(T_i, C_i)$, $\delta_i = I\{T_i \leq C_i\}$, and C_i is the right censoring variable which is independent of $(T_i, \boldsymbol{Z}_i)$.

1.2.2 Cox Model

As a special case of PH model (1.2.18), the Cox model (Cox, 1972) assumes a parametric form of $\psi(\boldsymbol{z})$:

$$\psi(\boldsymbol{z}, \boldsymbol{\beta}_0) = e^{\boldsymbol{\beta}_0^\top \boldsymbol{z}}, \tag{1.2.21}$$

where $\boldsymbol{\beta}_0 = (\beta_1, \cdots, \beta_k)^\top$ is a vector of regression parameters. Thus, the Cox model is written as

$$h(t; \boldsymbol{z}) = h_0(t) e^{\boldsymbol{\beta}_0^\top \boldsymbol{z}}. \tag{1.2.22}$$

Consider the two-sample case with right censoring (see (1.2.20) and Section 1.3.1 for the definition). In this case, $Z = 0$ or 1 as in (1.2.9), $\beta_0 \in \mathbb{R}$, and the data observed are as (1.2.20) with scalar Z_i's:

$$(V_i, \delta_i, Z_i), \qquad i = 1, \cdots, n. \tag{1.2.23}$$

Then, Cox model (1.2.22) is the same as PH model (1.2.10) with parameter $\psi(z) = e^{\beta_0 z}$:

$$h(t; z) = h_0(t) e^{\beta_0 z}. \tag{1.2.24}$$

Here, we see that when $Z = 1$, we have $h_{T1}(t) = h(t; 1) = h_0(t) e^{\beta_0}$; and when $Z = 0$, we have $h_{T0}(t) = h_0(t) = h(t; 0)$. Thus, (1.2.22) is the same as (1.2.7) with parameter $\psi = e^{\beta_0}$:

$$h_{T1}(t) = h_{T0}(t) e^{\beta_0}. \tag{1.2.25}$$

Note that as discussed earlier, we know that if $\psi = e^{\beta_0} < 1$ (if and only if $\beta_0 < 0$) then the treatment is effective; and if $\psi = e^{\beta_0} > 1$ (if and only if $\beta_0 > 0$) then the treatment is not effective. Thus, to check whether the treatment is effective using Cox model (1.2.25), we test

$$H_0 : \beta_0 = 0 \quad \text{vs.} \quad H_1 : \beta_0 < 0. \tag{1.2.26}$$

Next, we give a brief review on a simple method on the goodness-of-fit for Cox model (1.2.25), then discuss the estimation and tests on β_0 in (1.2.25).

<u>Goodness-of-Fit of Cox Model (1.2.25) with Data (1.2.23)</u>

The goodness-of-fit (GOF) of two-sample Cox model (1.2.25) can be tested using informal graphical methods. Let $\bar{F}_{T1}(t)$ and $\bar{F}_{T0}(t)$ be the survivor functions corresponding to $h_{T1}(t)$ and $h_{T0}(t)$, respectively. From Cox and Oakes (1984; page 70), we know that Cox model (1.2.25) is equivalent to

$$\bar{F}_{T1}(t) = \big(\bar{F}_{T0}(t)\big)^{e^{\beta_0}} \quad \Longleftrightarrow \quad -\ln \bar{F}_{T1}(t) = -e^{\beta_0} \ln \bar{F}_{T0}(t) \tag{1.2.27}$$

$$\Longleftrightarrow \quad \ln\big\{-\ln \bar{F}_{T1}(t)\big\} = \ln\big\{e^{\beta_0}\big[-\ln \bar{F}_{T0}(t)\big]\big\} = \beta_0 + \ln\big\{-\ln \bar{F}_{T0}(t)\big\}. \tag{1.2.28}$$

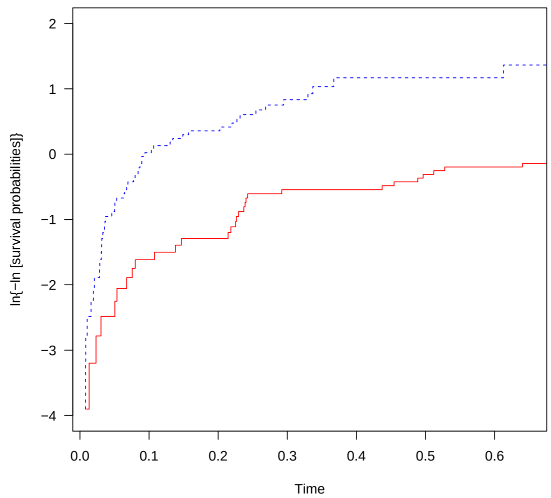
Therefore, to check the GOF of Cox model (1.2.25) with right censored data (1.2.23), we may examine the graphs for $\ln\{-\ln \hat{\bar{F}}_{T0}(t)\}$ and $\ln\{-\ln \hat{\bar{F}}_{T1}(t)\}$, where $\hat{F}_{T0}(t)$ and $\hat{F}_{T1}(t)$ are the Kaplan-Meier (KM) estimators (Kaplan and Meier, 1958) for $F_{T0}(t)$ and $F_{T1}(t)$, respectively.

<u>Simulation Study for (1.2.27)-(1.2.28)</u>

Consider $F_{T0} = \text{Exp}(1)$ in Cox model (1.2.25), where $\text{Exp}(\mu)$ represents an exponential distribution with mean μ. Then, we have $\bar{F}_{T0}(t) = e^{-t}$ for $t > 0$, and we have from (1.2.27)

$$\bar{F}_{T1}(t) = \big(e^{-t}\big)^{e^{\beta_0}} = e^{-e^{\beta_0} t}, \qquad \text{for } t > 0, \tag{1.2.29}$$

which means $F_{T1} = \text{Exp}(-e^{\beta_0})$. With this choice of F_{T0} and F_{T1}, we see that F_{T0}, F_{T1}, and



The solid line is $\ln\{-\ln \bar{F}_{n,T0}(t)\}$ for the control group
The dashed line is $\ln\{-\ln \bar{F}_{n,T1}(t)\}$ for the treatment group

Figure 1.1: Comparing $\ln\{-\ln \bar{F}_{n,T0}(t)\}$ and $\ln\{-\ln \bar{F}_{n,T1}(t)\}$ with $n = 50$

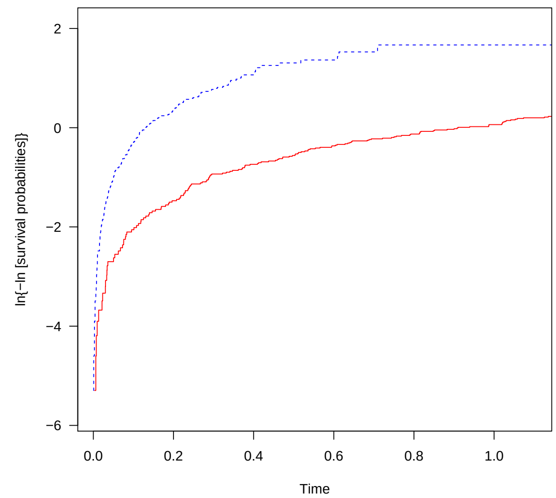
β_0 satisfy (1.2.27), thus they satisfy Cox model assumption (1.2.25).

Here, we consider the case when $\beta_0 = 2$. We generate sample $T_1, \cdots, T_{n_0}$ from $F_{T0} = \text{Exp}(1)$ and sample $X_1, \cdots, X_{n_1}$ from $F_{T1} = \text{Exp}(e^{-2})$, respectively, with $n = n_0 = n_1 = 50$, and then calculate the empirical d.f.'s $F_{n,T0}(t)$ and $F_{n,T1}(t)$ for $f_{T0}(t)$ and $f_{T1}(t)$, respectively. Figure 1.1 displays the comparison between $\ln\{-\ln \bar{F}_{n,T0}(t)\}$ and $\ln\{-\ln \bar{F}_{n,T1}(t)\}$ for $\beta_0 = 2$ and $n = 50$, where the portion of the curves displayed is on the intersection of their ranges,

$$\begin{aligned}[T_{(1)}, T_{(n)}] \cap [X_{(1)}, X_{(n)}] &= [0.00968113, 3.43680386] \cap [0.00804576, 0.70925274] \\ &= [0.00968113, 0.70925274].\end{aligned}$$

It is evident that Figure 1.1 shows that the difference between $\ln\{-\ln \bar{F}_{n,T0}(t)\}$ and $\ln\{-\ln \bar{F}_{n,T1}(t)\}$ is approximately 2 on the interval for $t > 0.05$, which is in agreement with the theoretical results in (1.2.27)-(1.2.28).

Also, this simulation study is repeated for case $n = n_0 = n_1 = 200$. Figure 1.2 displays the comparison between $\ln\{-\ln \bar{F}_{n,T0}(t)\}$ and $\ln\{-\ln \bar{F}_{n,T1}(t)\}$ for $\beta_0 = 2$ and $n = 200$,



The solid line is $\ln\{-\ln \bar{F}_{n,T0}(t)\}$ for the control group
The dashed line is $\ln\{-\ln \bar{F}_{n,T1}(t)\}$ for the treatment group

Figure 1.2: Comparing $\ln\{-\ln \bar{F}_{n,T0}(t)\}$ and $\ln\{-\ln \bar{F}_{n,T1}(t)\}$ with $n = 200$

where the portion of the curves displayed is on the intersection of their ranges,

$$\begin{aligned}[T_{(1)}, T_{(n)}] \cap [X_{(1)}, X_{(n)}] &= [0.00324821, 5.21481632] \cap [0.00087197, 1.14824665] \\ &= [0.00324821, 1.14824665].\end{aligned}$$

From Figure 1.2 it is evident that the difference between $\ln\{-\ln \bar{F}_{n,T0}(t)\}$ and $\ln\{-\ln \bar{F}_{n,T1}(t)\}$ is approximately 2 on the interval for $t > 0.05$. Note that the approximation in Figure 1.2 is better than that in Figure 1.1, which is expected as the sample size is larger for Figure 1.2.

Estimation and Tests on β_0 in (1.2.25) with Data (1.2.23)

Cox's partial likelihood estimate $\hat{\beta}_c$ for β_0 can be computed based on data (1.2.23) by maximizing the partial likelihood function given by Cox (1972), i.e. by solving the following estimating equation given by Tsiatis (1981):

$$n^{-1} \sum_{i=1}^{n} \delta_i Z_i - n^{-1} \sum_{i=1}^{n} \delta_i \frac{\sum_{j=1}^{n} Z_j e^{\beta Z_j} I\{V_j \geq V_i\}}{\sum_{j=1}^{n} e^{\beta Z_j} I\{V_j \geq V_i\}} = 0. \tag{1.2.30}$$

From Andersen and Gill (1982), we know

$$\sqrt{n}(\hat{\beta}_c - \beta_0) \xrightarrow{\mathcal{D}} N(0, \sigma^2), \quad \text{as } n \to \infty, \tag{1.2.31}$$

where $0 < \sigma^2 < \infty$ is a constant, which implies

$$\frac{\hat{\beta}_c - \beta_0}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{D}} N(0, 1), \quad \text{as } n \to \infty. \tag{1.2.32}$$

Note that we may use the bootstrap method to estimate the standard error of $\hat{\beta}_c$. As in Efron and Tibshirani (1986), draw B independent bootstrap samples $(V_1^{*b}, \delta_1^{*b}, Z_1^{*b}), \cdots,$ $(V_n^{*b}, \delta_n^{*b}, Z_n^{*b})$, $1 \leq b \leq B$, with replacement from observed data (1.2.23). For each bootstrap sample, compute Cox's partial likelihood estimate $\hat{\beta}_c^*(b)$ via estimating equation (1.2.30). The bootstrap estimate for the standard error of $\hat{\beta}_c$ is the sample standard deviation of $\hat{\beta}_c^*(1), \cdots, \hat{\beta}_c^*(B)$, which is given by

$$\widehat{se}_B = \left(\frac{\sum_{b=1}^{B} (\hat{\beta}_c^*(b) - \bar{\hat{\beta}}^*)^2}{B - 1} \right)^{1/2}, \tag{1.2.33}$$

where

$$\bar{\hat{\beta}}^* = B^{-1} \sum_{b=1}^{B} \hat{\beta}_c^*(b). \tag{1.2.34}$$

Then, the test statistic based on (1.2.32) for hypothesis test (1.2.26) is given by

$$\tau = \frac{\hat{\beta}_c}{\widehat{se}_B} \overset{\mathcal{D}}{\approx} N(0, 1), \quad \text{as } n \to \infty, \tag{1.2.35}$$

and we reject H_0 in favor of H_1 if $\tau \leq -z_\alpha$, where α is the significance level and z_α satisfies $P\{Z_0 \leq -z_\alpha\} = \alpha$ with Z_0 as the standard normal random variable. Moreover, the

$(1-\alpha)100\%$ confidence interval for β_0 is given by

$$\Big(\hat{\beta}_c - z_{\alpha/2}(\widehat{se}_B),\ \hat{\beta}_c + z_{\alpha/2}(\widehat{se}_B)\Big). \tag{1.2.36}$$

Next, we give an example discussed in Cox (1972) and Gehan (1965) and conduct a graphical check of the goodness-of-fit of the Cox model.

<u>Example 1.1</u>

As an example, consider a two-sample study discussed in Cox (1972) and Gehan (1965) on the maintenance of remissions in acute leukemia patients. In this study, T is the length of remission of a patient (in weeks), and each patient is either given a treatment (6-Mercaptopurine) or a control (placebo). One year after the start of the study, the following lengths of remissions were recorded:

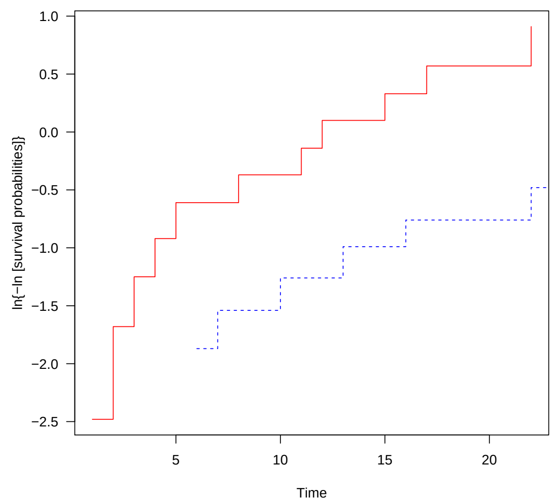
Table 1.1 Lengths of Remission of Leukemia Patients (weeks)

Treatment	6*	6	6	6	7	9*	10*	10	11*	13	16
	17*	19*	20*	22	23	25*	32*	32*	34*	35*	
Control	1	1	2	2	3	4	4	5	5	8	8
	8	8	11	11	12	12	15	17	22	23	

* denotes a right censored observation (see Section 1.3.1 for the definition of right censoring)

The data in Table 1.1 can be written as (1.2.23) with $n = 42$, where $Z_i = 0$ represents the control group and $Z_i = 1$ represents the treatment group. Thus, under Cox model assumption (1.2.25), we may use hypothesis test (1.2.26) to see whether 6-Mercaptopurine is an effective treatment for Leukemia patients.

To check the GOF for Cox model assumption (1.2.25) with Example 1.1, we may compute the KM estimators $\hat{F}_{T0}(t)$ and $\hat{F}_{T1}(t)$, using the formula in Shorack and Wellner (1986; page 293), for the d.f.'s of the control group and the treatment group, respectively. Figure 1.3 displays the comparison between $\ln\{-\ln \bar{\hat{F}}_{T0}(t)\}$ and $\ln\{-\ln \bar{\hat{F}}_{T1}(t)\}$, and shows that the two curves are approximately parallel on the intersection of their supports. Based on (1.2.28),



The solid line is $\ln\{-\ln \hat{\bar{F}}_{T0}(t)\}$ for the control group
The dashed line is $\ln\{-\ln \hat{\bar{F}}_{T1}(t)\}$ for the treatment group

Figure 1.3: Graphical Check of Goodness-of-Fit

we conclude that by the graphical method, the Leukemia data Table 1.1 fits Cox model assumption (1.2.25) reasonably well.

1.3 Censored Data

The term censoring (first used by Hald, 1949) is used to describe observations in a study which contain incomplete information. Although the concept of censoring came from biomedical research, censored observations occur in other areas of research such as social science, reliability, and economic research. The most frequently encountered type of censoring in practice is right censoring, thus analysis of right censored data has been a major area of research for statisticians over the past three decades. In recent years, attention has been focused on more complicated types of censored data, such as doubly censored data, interval censored data, and partly interval-censored data, due to their applications in important medical and epidemiological studies. For instance, doubly censored data were encountered in recent studies on breast cancer (Peer *et al.*, 1993; Ren and Gu, 1997; Ren and Peer, 2000), interval censored data were encountered in HIV/AIDS research (O'Brien *et al.*, 1994; De

Gruttola and Lagakos, 1989; Kim *et al.*, 1993; Ren, 2003), and partly interval-censored data were encountered in Huntington Disease studies (Cupples *et al.*, 1991) and Coronary Heart Disease studies (Odell *et al.*, 1992).

In Sections 1.3.1-1.3.4, we give the definitions of right censored data, doubly censored data, interval censored data, and partly interval-censored data, respectively, and for each type of censored data, we discuss some real data examples and review relevant asymptotic results. In Section 1.3.5, we review some recent results on the Cox model with censored data. As notations used throughout this dissertation, we let

$$T_1, \cdots, T_n \tag{1.3.1}$$

be an independently and identically distributed (iid) random sample from a continuous and nonnegative d.f. $F_T(t)$.

1.3.1 Right Censored Data

The observed data for sample (1.3.1) are:

$$\boldsymbol{O}_i = (V_i, \delta_i) \qquad 1 \leq i \leq n, \tag{1.3.2}$$

with

$$V_i = \begin{cases} T_i & \text{if} \quad T_i \leq C_i \qquad (\delta_i = 1) \\ C_i & \text{if} \quad T_i > C_i \qquad (\delta_i = 0), \end{cases} \tag{1.3.3}$$

where C_i is the right censoring variable and is independent from T_i.

Example: *Laryngeal Cancer Data*

Right censored data were encountered in a study on the incidence of death in male laryngeal cancer patients (Kardaun, 1983). The study was conducted at a minor hospital in The Netherlands during the period from 1970 to 1981. In the study, 90 male patients were diagnosed and treated for cancer of the larynx. For each patient, the age at the time of diagnosis (ranging from 41 to 86 years), the year of diagnosis, and the stage of the cancer were recorded. The stages of cancer were based on the T.N.M. classification used by the American Joint Committee for Cancer Staging. Here, the time T_i from the first treatment of the laryngeal cancer to the death of the patient due to the cancer is of interest. However, not all the patients died before the end of the study (March 1, 1981), thus T_i was not observed for every patient. Among the 90 males in the study group, 40 were still alive by the end of the study, yielding 40 right censored observations ($\delta_i = 0$), while 50 died from the cancer before the end of the study, yielding 50 uncensored observations ($\delta_i = 1$). Hence, this data set is a right censored sample (1.3.2) with $n = 90$.

Likelihood Function and Asymptotic Results

The likelihood function for $F_T(t)$ with right censored data (1.3.2) is given by

$$L(F) = \prod_{i=1}^{n} [1 - F(V_i)]^{1-\delta_i} [F(V_i) - F(V_i-)]^{\delta_i}, \tag{1.3.4}$$

where $F(t)$ is any distribution function. The *nonparametric maximum likelihood estimator* (NPMLE) for $F_T(t)$ is the function $\hat{F}_n(t)$ that maximizes this likelihood function. The product-limit estimator (Kaplan and Meier, 1958) has been shown to be the NPMLE $\hat{F}_n(t)$ for $F_T(t)$, and is given by

$$1 - \hat{F}_n(t) = \prod_{V_{(i)} \le t} \left\{1 - \frac{1}{n-i+1}\right\}^{\delta_{(i)}} = \prod_{V_{(i)} \le t} \left\{1 - \frac{\delta_{(i)}}{n-i+1}\right\}, \tag{1.3.5}$$

where $V_{(1)} \le \cdots \le V_{(n)}$, and $\delta_{(i)}$ is the corresponding δ for $V_{(i)}$ (Shorack and Wellner, 1986). Note that if there are ties in the V_i's, then the V_i that is uncensored ($\delta_i = 1$) is ranked ahead of the V_j that is censored ($\delta_j = 0$). Stute and Wang (1993) showed that, under certain conditions, we have $||\hat{F}_n - F_T|| \stackrel{a.s.}{\to} 0$, as $n \to \infty$. It was also shown that $\sqrt{n}(\hat{F}_n - F_T)$ weakly converges to a centered Gaussian process (Gill, 1983).

1.3.2 Doubly Censored Data

The observed data for sample (1.3.1) are:

$$\boldsymbol{O}_i = (V_i, \delta_i) \qquad 1 \le i \le n, \tag{1.3.6}$$

with

$$V_i = \begin{cases} T_i & \text{if} \quad D_i < T_i \le C_i \qquad (\delta_i = 1) \\ C_i & \text{if} \quad T_i > C_i \qquad\qquad (\delta_i = 2) \\ D_i & \text{if} \quad T_i \le D_i \qquad\qquad (\delta_i = 3), \end{cases} \tag{1.3.7}$$

where C_i and D_i are right and left censoring variables, respectively, with $P\{D_i < C_i\} = 1$, and (C_i, D_i) is independent from T_i.

Example: *Screening Mammograms Data*

Doubly censored data were found in a study concerned with the effectiveness of screening mammograms (Peer *et al.*, 1993; Ren and Gu, 1997; Ren and Peer, 2000). During the period from 1981 to 1990, nearly 30,000 women in Nijmegen, a city in The Netherlands of about 150,000 inhabitants, were invited for biennial screening mammograms, and 236 women with ages ranging from 41 to 84 years were diagnosed with breast cancer. In this study, the age T_i when a tumor volume developed was of interest. Among the 236 women, 45 had tumor volumes observed at the first screening mammogram; yielding 45 left censored observations, while 79 did not have tumor volumes observed at the last available screening mammogram; yielding 79 right censored observations. The rest of the 112 women were observed to have tumor growth at the screening mammograms; yielding 112 uncensored observations. Thus, this data set is a doubly censored sample (1.3.6) with $n = 236$. Other examples of doubly censored data can be found in Chang and Yang (1987), Mykland and Ren (1996), Ren and Peer (2000), among others.

Likelihood Function and Asymptotic Results

The likelihood function for $F_T(t)$ with doubly censored data (1.3.6) is given in Mykland

and Ren (1996). The NPMLE for $F_T(t)$ is the function $\hat{F}_n(t)$ that maximizes this likelihood function. An algorithm to compute $\hat{F}_n(t)$ is given in Mykland and Ren (1996). Gu and Zhang (1993) showed that, under certain conditions, we have $||\hat{F}_n - F_T|| \stackrel{a.s.}{\to} 0$, as $n \to \infty$. It was also shown that $\sqrt{n}(\hat{F}_n - F_T)$ weakly converges to a centered Gaussian process (Gu and Zhang, 1993).

1.3.3 Interval Censored Data

Interval Censored Case 1 Data

The observed data for sample (1.3.1) are:

$$\boldsymbol{O}_i = (C_i, \delta_i) \qquad 1 \le i \le n, \tag{1.3.8}$$

where $\delta_i = I\{T_i \le C_i\}$, and C_i is independent from T_i.

Interval Censored Case 2 Data

The observed data for sample (1.3.1) are:

$$\boldsymbol{O}_i = (C_i, D_i, \delta_i) \qquad 1 \le i \le n, \tag{1.3.9}$$

where

$$\delta_i = \begin{cases} 1 & \text{if} \quad D_i < T_i \le C_i \\ 2 & \text{if} \quad T_i > C_i \\ 3 & \text{if} \quad T_i \le D_i, \end{cases} \tag{1.3.10}$$

where (C_i, D_i) is independent from T_i, and $P\{D_i < C_i\} = 1$.

Example 1: *HIV Transmission Data*

From 1987 to 1992, at study sites in California, New Jersey, and New York City, a retrospective study was conducted on the incidence of transmission of HIV from male blood transfusion patients to their female sex partners (O'Brien *et al.*, 1994). All the male patients contracted HIV due to an infected blood transfusion at a known date sometime after 1978. Demographic and medical background information was obtained for the male and female of each couple during a separate interview. Additionally, if the female had not already been diagnosed with HIV, she was tested for infection of HIV at the time of the interview. Here, the time T_i between the infection of the male partner and the contraction of HIV by the female partner was of interest. However, only the time C_i between the infection of the male partner and the interview was observed (in months) with an indicator function δ_i, which indicates whether the female contracted HIV before the interview. The study group consisted of 32 males aged 18 years and older and the 32 female sex partners. Of the 32 females, 7 contracted HIV before the time of the interview ($\delta_i = 1$), while the remaining 25 had not contracted HIV by the time of the interview ($\delta_i = 0$). Clearly, in this example C_i is independent of T_i, thus this data set is an interval censored Case 1 sample (1.3.8) with $n = 32$.

Example 2: *HIV Infection Data*

During the period from 1978 to 1988, 262 individuals with either Type A or Type B hemophilia were treated at Hôpital Kremlin Bicêtre and Hôpital Cœur des Yvelines in France (De Gruttola and Lagakos, 1989; Kim *et al.*, 1993; Ren, 2003). Each of the patients had blood samples taken and stored at one of the hospitals, and these samples were later tested for infection of HIV. All the infected individuals were believed to have become infected due to contaminated blood factor they received for their hemophilia. This study is concerned with the time T_i of infection of HIV, measured in 6-month intervals. However, the only available information for each individual is that $T_i \in [D_i, C_i]$, where $D_i < C_i$. Note that the individuals infected with HIV at entry are assigned $D_i = 1$, which denotes July 1, 1978. For example, [1,4] denotes an individual who was found to be infected with HIV at entry during the fourth 6-month interval. Among the 262 hemophilia patients in the study group, 25 were found to be infected with HIV at entry ($T_i \in [1, C_i]$, $\delta_i = 3$), 40 never became infected with HIV before the last test ($T_i \in [D_i, \infty)$, $\delta_i = 2$), while 197 became infected with HIV between two tests ($T_i \in [D_i, C_i]$, $\delta_i = 1$). Note that clearly (C_i, D_i) is independent of T_i since the blood samples were stored without intentions of doing a test for HIV. Therefore, this data set is an interval censored Case 2 sample (1.3.9) with $n = 262$.

<u>Likelihood Function and Asymptotic Results</u>

The likelihood functions for $F_T(t)$ with both interval censored Case 1 data (1.3.8) and interval censored Case 2 data (1.3.9) are given in Groeneboom and Wellner (1992). For each case, the NPMLE for $F_T(t)$ is the function $\hat{F}_n(t)$ that maximizes the corresponding likelihood function. A method to compute $\hat{F}_n(t)$ with interval censored data is given in Groeneboom and Wellner (1992). Groeneboom and Wellner (1992) showed that, under certain conditions, $||\hat{F}_n - F_T|| \overset{a.s.}{\to} 0$, as $n \to \infty$. It was also shown that, under certain conditions $\hat{F}_n(t)$ with interval censored Case 1 data has $n^{1/3}$ rate of convergence (Groeneboom and Wellner, 1992). That is, for any fixed point t_0,

$$n^{1/3}\{\hat{F}_n(t_0) - F_T(t)(t_0)\} \overset{\mathcal{D}}{\to} c_0 Z, \qquad as\ n \to \infty, \tag{1.3.11}$$

where c_0 is a constant, and $Z = \arg\min_t \left\{W(t) + t^2\right\}$ with W as the standard two-sided Brownian motion. For interval censored Case 2 data, this result is unknown.

1.3.4 Partly Interval-Censored Data

<u>'Case 1' Partly Interval-Censored Data</u>

The observed data for sample (1.3.1) are:

$$\boldsymbol{O}_i = \begin{cases} T_i & \text{if} \quad 1 \le i \le k_0 \\ (C_i, \delta_i) & \text{if} \quad k_0 + 1 \le i \le n, \end{cases} \tag{1.3.12}$$

where C_i is independent from T_i and $\delta_i = I\{T_i \le C_i\}$.

General Partly Interval-Censored Data

The observed data for sample (1.3.1) are:

$$\boldsymbol{O}_i = \begin{cases} T_i & \text{if} \quad 1 \leq i \leq k_0 \\ (\boldsymbol{C}, \boldsymbol{\delta}_i) & \text{if} \quad k_0 + 1 \leq i \leq n, \end{cases} \tag{1.3.13}$$

where, for N potential examination times $C_1 < \cdots < C_N$, $\boldsymbol{C} = (C_1, \cdots, C_N)$ with $C_0 = 0$ and $C_{N+1} = \infty$, and $\boldsymbol{\delta}_i = \left(\delta_i^{(1)}, \cdots, \delta_i^{(N+1)}\right)$ with $\delta_i^{(j)} = I\{C_{j-1} < T_i \leq C_j\}$ for $j = 1, \cdots, N+1$. In other words, T_i is either known exactly or is known to fall in one of the intervals $(0, C_1]$, $(C_1, C_2], \cdots, (C_N, \infty)$.

Example 1: *Huntington Disease Data*

During the period from 1980 to 1987, the Huntington Disease Center in Boston, MA collected information on pedigrees that have a history of Huntington Disease (HD). A study was later conducted on the incidence of HD in these pedigrees (Cupples *et al.*, 1991). A proband is defined here as the first person of a pedigree to contact the Huntington Disease Center. Not all of the probands were affected with HD. This study considers those pedigrees for whom the proband was affected with HD. The time T_i of development of HD is of interest. However, for some individuals in the study the exact time of development of HD is unknown, but C_i was observed on the individual, where C_i is either the current age or the age at death for the patient or the age before which T_i occurred. For some individuals, neither T_i nor C_i was observed. Out of the total 1,364 individuals, here we only consider those 965 individuals with information either T_i or C_i. Among these 965 individuals, 76 of them had exact ages at onset of HD observed, yielding 76 T_i's; 80 were affected with HD before C_i, yielding 80 of $(C_i, 1)$'s; 809 were unaffected with HD by time C_i, yielding 809 of $(C_i, 0)$'s. Therefore, the data set of 965 individuals is a 'Case 1' partly interval-censored sample (1.3.12) with $n = 965$.

Example 2: *Coronary Heart Disease Data*

Since 1971, offspring from the original Framingham Heart Study cohort in Framingham, MA and their spouses have been identified and studied (Feinleib *et al.*, 1975; Odell *et al.*, 1992). The women were given three clinical examinations to study the age at onset of angina pectoris (AP). The time between the first and second exam was approximately 8 years, and the time between the second and third exam was approximately 4 years. One of the interests of this study is the time T_i of the first occurrence of AP. Actual dates of the first occurrence of AP were recorded if available. However, for some individuals in the study, the time of first occurrence is only known to be between two of the clinical exams. So only an interval could be recorded for these individuals. Of the 2,568 women in the study free of AP at the time of the first exam, 16 had the first interval recorded ($T_i \in (C_1, C_2]$), 13 had the second interval recorded ($T_i \in (C_2, C_3]$), 2,531 did not develop AP by the time of the last exam ($T_i \in (C_3, \infty)$), and 8 had exact times of first occurrence of AP recorded. Hence, this data set is a general partly interval-censored sample (1.3.13) with $n = 2568$.

Likelihood Function and Asymptotic Results

The likelihood functions for $F_T(t)$ with both 'Case 1' partly interval-censored data (1.3.12) and general partly interval-censored data (1.3.13) are given in Huang (1999). For each case, the NPMLE for $F_T(t)$ is the function $\hat{F}_n(t)$ that maximizes the corresponding likelihood function. A method to compute $\hat{F}_n(t)$ with partly interval-censored data is given in Huang (1999). Huang (1999) showed that, under certain conditions, $||\hat{F}_n - F_T|| \xrightarrow{a.s.} 0$, as $n \to \infty$. It was also shown that for both 'Case 1' and general partly interval-censored data $\sqrt{n}(\hat{F}_n - F_T)$ weakly converges to a centered Gaussian process (Huang, 1999).

1.3.5 Cox Model with Various Types of Censored Data

Over the past three decades, extensive research has been done for the Cox model with non-censored and right censored data (1.3.2), and some research has also been done to deal with interval censored data (1.3.8)-(1.3.9). As follows, we review some works on Cox model (1.2.24) with right censored data (1.3.2) and interval censored data (1.3.8)-(1.3.9).

Kalbfleisch and Prentice (1973) studied the Cox model with right censored data (1.3.2) and used the marginal likelihood to estimate the regression parameters without having to estimate the baseline hazard function.

Efron (1977) and Cox and Oakes (1984) studied Cox's partial likelihood estimate $\hat{\beta}_C$ for β_0 for the Cox model with right censored data (1.3.2). As mentioned in Section 1.2.2, Cox's partial likelihood estimate $\hat{\beta}_C$ can be computed for right censored data (1.3.2) by solving estimating equation (1.2.30). Efron (1977) showed that $\hat{\beta}_C$ is asymptotically efficient, while Cox and Oakes (1984; page 123) pointed out that the loss in precision from using the partial likelihood can be rather substantial.

Ren and Zhou (2011) also studied the Cox model with right censored data (1.3.2). They used the empirical likelihood approach to profile out nuisance parameter $F_{T0}(t)$ to obtain the full-profile likelihood function for β_0 and the maximum likelihood estimator (MLE) for (β_0, F_{T0}). Their simulation studies show that the MLE has small-sample advantage over Cox's partial likelihood estimator $\hat{\beta}_C$.

Satten (1996) studied the Cox model with interval censored Case 2 data (1.3.9) and developed a method that extends the marginal likelihood approach of Kalbfleisch and Prentice (1973) so that it can be used with this type of data. Using this method, Satten showed that the regression parameters can be calculated for the Cox model with interval censored Case 2 data (1.3.9) without estimating the baseline hazard function.

Huang (1996) studied the Cox model with interval censored Case 1 data (1.3.8) and found that the MLE for the regression parameters is asymptotically normal and efficient with $\sqrt{n}$ convergence rate, even though the nonparametric MLE for the baseline integrated hazard function has only a $n^{1/3}$ convergence rate.

Pan (1999) studied the Cox model with interval censored Case 2 data (1.3.9) and extended the iterative convex minorant algorithm, given in Groeneboom and Wellner (1992), to compute the MLE for the regression parameters.

Farrington (2000) extended the Cox-Snell residuals (Cox and Snell, 1968), Lagakos residuals (Lagakos, 1980), deviance residuals (Therneau, Grambsch, and Fleming, 1990), and

Schoenfeld residuals (Schoenfeld, 1982) so they may be used under Cox model (1.2.25) with interval censored Case 2 data (1.3.9). In particular, the Cox-Snell residuals can be used to detect non-proportional hazards, while the Lagakos residuals are used to check regression relationships.

It is important to note that, although the papers by Satten (1996), Huang (1996), Pan (1999), and Farrington (2000) all use the same definitions for interval censoring, the terms used for censored data are not always consistent. For example, Sun, Liao and Pagano (1999) and Pan (2001) studied the Cox model with "doubly censored data," but after close analysis of these articles one can see that the failure time they considered is the time between two events in which the first event is interval censored and the second event is right censored. This type of censored data is not our doubly censored data (1.3.6).

One may also note that above mentioned papers together with earlier works on the Cox model, say Andersen *et al.* (1993), only deal with non-censored data, right censored data (1.3.2), and interval censored data (1.3.8)-(1.3.9). Currently, there has not been any published work on the Cox model with doubly censored data (1.3.6) or with partly interval-censored data (1.3.12)-(1.3.13).

In this dissertation, we develop general estimation methods for the Cox model with various types of censored data. The key to our approach is the weighted empirical likelihood function (Ren, 2001), which is reviewed in the next section.

1.4 Likelihood Inference

It is well-known that likelihood is probably the most important concept used for inference with parametric models as well as nonparametric models. Likelihood methods are preferred over other methods partly because the estimators usually possess desirable asymptotic properties such as consistency, efficiency and asymptotic normality.

The likelihood function is defined as the probability of observing the available data. This expression can be used to derive estimators and construct hypothesis tests and confidence sets even if the distribution of the data is unknown. When the distribution of the data is known, the parametric likelihood function can be used to find the maximum likelihood estimator for the parameter and construct hypothesis tests and confidence sets. To avoid specifying a distribution for the data, empirical likelihood (Owen, 1988), a nonparametric method, can be used to construct hypothesis tests and confidence sets. Empirical likelihood combines the reliability of nonparametric methods with the flexibility and effectiveness of the likelihood approach. However, when imposing a model assumption with complicated types of censored data, the empirical likelihood function can be very difficult or impossible to maximize. To deal with these issues, Ren (2001, 2008a) proposed weighted empirical likelihood, which provides a simple and direct way to incorporate some model assumptions in the derivation of the likelihood function for various types of censored data. So far, the results on weighted empirical likelihood have shown to be favorable to alternative methods.

In Sections 1.4.1-1.4.3, we describe parametric likelihood, empirical likelihood, and weighted empirical likelihood, respectively, and review some relevant inference problems.

1.4.1 Parametric Likelihood

Parametric likelihood is used to construct hypothesis tests and confidence sets for an unknown parameter when the underlying distribution is known.

Suppose $T_1, \cdots, T_n$ is a random sample from a distribution with known density function $f_T(t;\, \theta)$, where the parameter $\theta \in \Theta$ is unknown. Let $t_1, \cdots, t_n$ be the observed data and $\boldsymbol{T} = \boldsymbol{T}_\theta = (T_1, \cdots, T_n)$. Then

$$\begin{aligned} P\big\{\text{Observe the given data}\big\} &= P\{T_1 = t_1, \cdots, T_n = t_n\} \\ &\overset{\text{iid}}{=} \prod_{i=1}^{n} P\{T = t_i\} = \prod_{i=1}^{n} f_T(t_i;\, \theta). \end{aligned} \tag{1.4.1}$$

Thus, the likelihood function for θ based on random sample $T_1, \cdots, T_n$ is given by

$$L(\theta \,|\, \boldsymbol{T}) = \prod_{i=1}^{n} f_T(T_i;\, \theta). \tag{1.4.2}$$

The value of θ which maximizes $L(\theta \,|\, \boldsymbol{T})$ over the whole parameter space Θ is called the maximum likelihood estimator (MLE) and is denoted by $\hat{\theta}$.

Suppose we wish to test

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_0^c, \tag{1.4.3}$$

where Θ_0 is the subset of the parameter space under the null hypothesis. Then, the likelihood ratio test statistic for (1.4.3) is given by

$$R(\boldsymbol{T}; \theta) = \frac{\sup_{u \in \Theta_0} L(u \,|\, \boldsymbol{T})}{\sup_{u \in \Theta} L(u \,|\, \boldsymbol{T})} = \sup_{u \in \Theta_0} \frac{L(u \,|\, \boldsymbol{T})}{L(\hat{\theta} \,|\, \boldsymbol{T})}, \tag{1.4.4}$$

where $\hat{\theta}$ is the MLE for θ.

For the rest of this subsection, we consider the simpler test of hypothesis

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta \neq \theta_0. \tag{1.4.5}$$

Under test (1.4.5), the likelihood ratio test statistic in (1.4.4) becomes

$$R(\boldsymbol{T}; \theta) = \frac{L(\theta_0 \,|\, \boldsymbol{T}_\theta)}{L(\hat{\theta} \,|\, \boldsymbol{T}_\theta)}. \tag{1.4.6}$$

To obtain the rejection region for hypothesis test (1.4.5), note that $\hat{\theta} \approx \theta$ because $\hat{\theta}$ is a consistent estimator for θ. Now, if H_0 holds in (1.4.5), then $\hat{\theta}$ will be close to θ_0, thus $R(\boldsymbol{T}; \theta)$ in (1.4.5) should be close to 1. On the other hand, if H_0 does not hold, then $\hat{\theta}$ will be significantly different from θ_0 because $\theta \neq \theta_0$, thus $R(\boldsymbol{T}; \theta)$ should be small since $\hat{\theta}$ is the MLE. Therefore, the rejection region for (1.4.5) is $\{\boldsymbol{T} \mid R(\boldsymbol{T}; \theta) \leq c\}$ for some predetermined constant $0 < c < 1$.

Now, let

$$r_\theta(t) = \sup_{u=t} \frac{L(u \mid \boldsymbol{T}_\theta)}{L(\hat{\theta} \mid \boldsymbol{T}_\theta)} = \frac{L(t \mid \boldsymbol{T}_\theta)}{L(\hat{\theta} \mid \boldsymbol{T}_\theta)}. \tag{1.4.7}$$

Wilks (1938) showed that $-2\log r_{\theta_0}(\theta_0)$ has a limiting chi-squared distribution, where $r_{\theta_0}(\theta_0)$ is $R(\boldsymbol{T};\theta)$ in (1.4.6) under the null hypothesis. Therefore,

$$\begin{aligned} P\{\text{Type I error}\} &= P\{\text{reject } H_0 \mid H_0 \text{ is true}\} = P\{R(\boldsymbol{T};\theta) \leq c \mid \theta = \theta_0\} \\ &= P\{r_{\theta_0}(\theta_0) \leq c\} = P\{-2\log r_{\theta_0}(\theta_0) \geq -2\log c\} \\ &\overset{\text{Wilks}}{\approx} P\{\chi^2_{(1)} \geq -2\log c\}, \end{aligned} \tag{1.4.8}$$

where $\chi^2_{(1)}$ is a chi-squared random variable with 1 degree of freedom. For a given significance level $0 < \alpha < 1$, we can determine the value of c according to

$$\alpha = P\{\chi^2_{(1)} \geq -2\log c\}. \tag{1.4.9}$$

Note that the acceptance region for hypothesis test (1.4.5) is

$$A(\theta_0) = \{\boldsymbol{T} \mid R(\boldsymbol{T};\theta) \geq c\} = \left\{ \boldsymbol{T} \,\middle|\, \frac{L(\theta_0 \mid \boldsymbol{T}_\theta)}{L(\hat{\theta} \mid \boldsymbol{T}_\theta)} \geq c \right\}. \tag{1.4.10}$$

Thus, the $(1-\alpha)100\%$ confidence set for $\theta = \theta_0$ is given by

$$C(\boldsymbol{T}) = \left\{ t \,\middle|\, \frac{L(t \mid \boldsymbol{T}_{\theta_0})}{L(\hat{\theta} \mid \boldsymbol{T}_{\theta_0})} \geq c \right\} = \{t \mid r_{\theta_0}(t) \geq c\}. \tag{1.4.11}$$

To see that (1.4.11) is the $(1-\alpha)100\%$ confidence set for θ_0, note that

$$\begin{aligned} P\{\theta_0 \in C(\boldsymbol{T})\} &= P\{r_{\theta_0}(\theta_0) \geq c\} = P\{-2\log r_{\theta_0}(\theta_0) \leq -2\log c\} \\ &\overset{\text{Wilks}}{\approx} P\{\chi^2_{(1)} \leq -2\log c\} \overset{(1.4.9)}{=} 1-\alpha. \end{aligned} \tag{1.4.12}$$

Note that the inference method described here is a parametric method, which assumes an explicit distribution for the data. Such methods are potentially much more powerful than nonparametric methods, but only if the distribution assumption used is correct. Using this method, the MLE is is shown to be usually consistent, asymptotically normal and efficient (Casella and Berger, 2002), and Wilks' theorem provides a way to construct confidence sets for the parameter. However, in practical situations, the underlying distribution is often unknown and assuming an incorrect distribution can be detrimental to confidence sets and hypothesis tests. In this situation, empirical likelihood (Owen, 1988) may be used to construct confidence sets and hypothesis tests without assuming a distribution for the data.

1.4.2 Empirical Likelihood

Empirical likelihood is analogous to parametric likelihood and is described as follows. Consider random sample (1.3.1) and let $\boldsymbol{T} = (T_1, \cdots, T_n)$. Owen (1988) defined the *empirical likelihood function* as

$$L(F) = \prod_{i=1}^{n} \left[F(T_i) - F(T_i-)\right], \tag{1.4.13}$$

where F is any distribution function. The *empirical distribution function*

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} I\{T_i \leq t\}, \tag{1.4.14}$$

is well-known to be the NPMLE of $F_T(t)$ since it maximizes likelihood function (1.4.13) over all d.f.'s F. Note that F_n in (1.4.14) is usually consistent, asymptotically normal and efficient.

Often a parameter θ of a d.f. F can be expressed as a statistical functional: $\theta = T(F)$. For instance, the mean is given as $\theta = T(F) = \int t dF(t)$, and the median is given as $\theta = F^{-1}(\frac{1}{2})$. As for parametric likelihood inference, if we wish to test (1.4.3), for $\theta = T(F_T)$, the empirical likelihood ratio test statistic for (1.4.3) is given by

$$R(\boldsymbol{T}; F_T) = \frac{\sup_{T(F)\in\Theta_0} L(F)}{\sup_F L(F)} = \sup_{T(F)\in\Theta_0} \frac{L(F)}{L(F_n)}, \tag{1.4.15}$$

where $L(F)$ and F_n are given in (1.4.13) and (1.4.14), respectively.

For the rest of this subsection, we consider the simpler test (1.4.5). Here, the empirical likelihood ratio test statistic for (1.4.5) is given by

$$R(\boldsymbol{T}; F_T) = \sup_{T(F)=\theta_0} \frac{L(F)}{L(F_n)}. \tag{1.4.16}$$

To obtain the rejection region for hypothesis test (1.4.5), note that $F_n \approx F_T$ because F_n is a consistent estimator for F_T. Now, if H_0 holds in (1.4.5) (i.e. $T(F_T) = \theta_0$), then (1.4.16) becomes

$$1 \geq R(\boldsymbol{T}; F_T) = \frac{\sup_{T(F)=\theta_0} L(F)}{L(F_n)} \geq \frac{L(F_T)}{L(F_n)}. \tag{1.4.17}$$

Since $F_n \approx F_T$, we have $\frac{L(F_T)}{L(F_n)} \approx \frac{L(F_n)}{L(F_n)} = 1$, which means that $R(\boldsymbol{T}; F_T)$ should be close to 1. On the other hand, if H_0 does not hold (i.e. $T(F_T) \neq \theta_0$), then $T(F_n)$ is far from θ_0 because $F_n \approx F_T$ implies $T(F_n) \approx T(F_T)$. Thus, for any F satisfying $T(F) = \theta_0$, F is not close to F_n (otherwise we would have $\theta_0 = T(F) \approx T(F_n)$, a contradiction). Consequently, $\sup_{T(F)=\theta_0} L(F)$ is not close to $L(F_n)$ because F_n is the unique NPMLE. Hence, $R(\boldsymbol{T}; F_T)$ should be small (i.e. not close to 1) when H_0 does not hold. Therefore, the rejection region for (1.4.5) is $\{\boldsymbol{T} \mid R(\boldsymbol{T}; F_T) \leq c\}$ for some predetermined constant $0 < c < 1$.

Now, let

$$r(t) = \sup_{T(F)=t} \frac{L(F)}{L(F_n)}. \tag{1.4.18}$$

Owen (1988) proved the analog to Wilks' (1938) theorem. In particular, for the mean $\theta_0 = E(T) = \int t dF_T(t)$, Owen (1988) showed:

Theorem 1.1. *(Owen, 1988) Assume $\int |t|^3 dF_T(t) < \infty$. Then, under $H_0 : \theta = \theta_0$, we have $-2\log r(\theta_0) \xrightarrow{\mathcal{D}} \chi^2_{(1)}$, as $n \to \infty$.*

Therefore, we have

$$\begin{aligned} P\{\text{Type I error}\} &= P\{\text{reject } H_0 \mid H_0 \text{ is true}\} = P\{R(\boldsymbol{T}; F_T) \le c \mid \theta = \theta_0\} \\ &= P\{r(\theta_0) \le c\} = P\{-2\log r(\theta_0) \ge -2\log c\} \\ &\approx P\{\chi^2_{(1)} \ge -2\log c\}. \end{aligned} \tag{1.4.19}$$

For a given significance level $0 < \alpha < 1$, we can determine the value of c according to

$$\alpha = P\{\chi^2_{(1)} \ge -2\log c\}. \tag{1.4.20}$$

The acceptance region for hypothesis test (1.4.5) is

$$A(\theta_0) = \{\boldsymbol{T} \mid R(\boldsymbol{T}; F_T) \ge c\} = \left\{\boldsymbol{T} \,\middle|\, \sup_{T(F)=\theta_0} \frac{L(F)}{L(F_n)} \ge c\right\}. \tag{1.4.21}$$

Thus, the $(1-\alpha)100\%$ confidence set for $\theta = \theta_0 = \int t df_{T0}(t)$ is given by

$$C(\boldsymbol{T}) = \left\{t \,\middle|\, \sup_{T(F)=t} \frac{L(F)}{L(F_n)} \ge c\right\} = \{t \mid r(t) \ge c\} \tag{1.4.22}$$

$$= \left\{T(F) \,\middle|\, \frac{L(F)}{L(F_n)} \ge c\right\}, \tag{1.4.22a}$$

where in most cases, the equivalence between (1.4.22) and (1.4.22a) can be shown under certain conditions. Owen (1988) showed that (1.4.22) is an interval, $C(\boldsymbol{T}) = [T_L, T_U]$, where $T_L = \inf_{F \in E_c} \int t dF$ and $T_U = \sup_{F \in E_c} \int t dF$ with

$$E_c = \left\{ F \,\Big|\, \frac{L(F)}{L(F_n)} \geq c,\ F(t) = \sum_{i=1}^{n} p_i I\{T_i \leq t\},\ 0 \leq p_i \leq 1,\ \sum_{i=1}^{n} p_i = 1 \right\}.$$

To see that $C(\boldsymbol{T})$ is a $(1-\alpha)100\%$ confidence interval for $\theta_0 = \int t dF_T(t)$, note that

$$\begin{aligned} P\{\theta_0 \in C(\boldsymbol{T})\} &= P\{T_L \leq \theta_0 \leq T_U\} = P\{r(\theta_0) \geq c\} \\ &= P\{-2\log r(\theta_0) \leq -2\log c\} \overset{\text{Owen}}{\approx} P\{\chi^2_{(1)} \leq -2\log c\} \\ &\overset{(1.4.20)}{=} 1-\alpha. \end{aligned} \tag{1.4.23}$$

Note that empirical likelihood is preferred over other nonparametric methods partly because it has been shown that the empirical log-likelihood ratio usually has an asymptotic chi-squared distribution (Owen, 1988) and that inferences based on the empirical likelihood function are of comparable accuracy to alternative methods (see Owen, 1990, 1991; DiCiccio *et al.*, 1991; Qin and Lawless, 1994; Mykland, 1995; among others). In particular, DiCiccio, Hall and Romana (1991) showed that empirical likelihood is Bartlett-correctable. However, the empirical likelihood method is not ideal when imposing a model assumption with complicated types of censored data because the likelihood function can be very difficult or impossible to maximize. To deal with these issues, Ren (2001, 2008a) proposed weighted empirical likelihood, which provides a simple and direct way to incorporate some model assumptions in the derivation of the likelihood function for various types of censored data.

1.4.3 Weighted Empirical Likelihood

Consider random sample (1.3.1) and let $\{\boldsymbol{O}_i | \, 1 \leq i \leq n\}$ denote the observed censored data for random sample (1.3.1), where the censoring could be right censoring (1.3.2), doubly censoring (1.3.6), interval censoring (1.3.8)-(1.3.9), or partly-interval censoring (1.3.12)-

(1.3.13). Instead of studying the different types of censored data separately as in the empirical likelihood approach, Ren (2001) proposed a *weighted empirical likelihood function*, which is formulated in a unified form depending only on the probability mass of the NPMLE $\hat{F}_n(t)$ for $F_T(t)$. The weighted empirical likelihood by Ren (2001, 2008a) is given as follows.

As reviewed in Section 1.3, the likelihood function for each of the types of censored data above mentioned has been given in literature. The NPMLE $\hat{F}_n$ for F_T is the solution which maximizes the likelihood function and is shown to be a strong uniform consistent estimator of F_T under some suitable conditions (see Section 1.3 for details). Moreover, it has been shown that for observed censored data $\{\boldsymbol{O}_i |\, 1 \leq i \leq n\}$, there exist m distinct points $U_1 < \cdots < U_m$ along with $\hat{p}_j > 0$, $1 \leq j \leq m$, such that the NPMLE $\hat{F}_n$ can be expressed as

$$\hat{F}_n(t) = \sum_{j=1}^{m} \hat{p}_j I\{U_j \leq t\}; \tag{1.4.24}$$

see Kaplan and Meier (1958) for right censored data (1.3.2), Mykland and Ren (1996) for doubly censored data (1.3.6), Groeneboom and Wellner (1992) for interval censored data (1.3.8)-(1.3.9), and Huang (1999) for partly-interval censored data (1.3.12)-(1.3.13). For instance, with right censored data (1.3.2), Kaplan and Meier (1958) showed that the U_j's are just the the uncensored observations in (1.3.2). Since $\hat{F}_n$ is a strong uniform consistent estimator of F_T, we may expect a random sample $T_1^*, \cdots, T_n^*$ from $\hat{F}_n$ to behave asymptotically the same as $T_1, \cdots, T_n$. Let F_n^* denote the empirical d.f. of $T_1^*, \cdots, T_n^*$. Now, since $\hat{F}_n \approx F_n^*$, we have

$$\begin{aligned}\prod_{i=1}^{n} P\{T = T_i\} \approx \prod_{i=1}^{n} P\{T^* = T_i^*\} &= \prod_{j=1}^{m} P\{T^* = U_j\}^{n[F_n^*(U_j) - F_n^*(U_j-)]} \\ &\approx \prod_{j=1}^{m} P\{T^* = U_j\}^{n[\hat{F}(U_j) - \hat{F}(U_j-)]} = \prod_{j=1}^{m} P\{T^* = U_j\}^{n\hat{p}_j}.\end{aligned}$$

Thus, the weighted empirical likelihood function (Ren, 2001) is given by

$$\hat{L}(F) = \prod_{i=1}^{m} [F(U_i) - F(U_i-)]^{n\hat{p}_i}, \tag{1.4.25}$$

where F is any distribution function. The weighted empirical likelihood function (1.4.25) may be viewed as the asymptotic version of the empirical likelihood function $L(F)$ for censored data; see arguments in Ren (2008a). Note that when there is no censoring, the weighted empirical likelihood function (1.4.25) is the same as Owen's empirical likelihood function (1.4.13); see arguments in Ren (2001).

The results on weighted empirical likelihood have shown to be favorable to alternative methods. In Ren (2001), it is shown that the weighted empirical likelihood ratio confidence interval for the mean with various types of censored data has comparable coverage accuracy to alternative methods, including the nonparametric bootstrap-t. In Ren (2008a), it is shown that for general two-sample semiparametric models with various types of censored data, the weighted empirical likelihood-based semiparametric maximum likelihood estimator for the underlying parameter and distribution have desirable asymptotic properties. In Ren (2008b), smoothed weighted empirical likelihood ratio confidence intervals for quantiles are constructed in a unified framework for various types of censored data and the coverage accuracy equation for the weighted empirical likelihood confidence interval is derived, which generally guarantees at least '*first order*' accuracy. Simulation studies show (Ren, 2008b) that for right censored data (1.3.2), the smoothed weighted empirical likelihood ratio confidence intervals are generally shorter than existing empirical likelihood-based confidence intervals and provide comparable coverage accuracy. In addition, simulation studies also show (Ren 2008b) that for interval censored data (1.3.8)-(1.3.9), the smoothed weighted empirical likelihood confidence intervals perform favorably compared to existing methods.

But, so far, weighted empirical likelihood has only been applied to the univariate case. In this dissertation, we extend the weighted empirical likelihood method to the multivariate case which provides the tool to study the Cox model problem with data (1.2.19), where the survival time T is subject to any of the types of censoring discussed in Section 1.3.

1.5 Summary of this Dissertation's Results

The main results of this dissertation are organized as follows. In Chapter 2, we derive the bivariate nonparametric maximum likelihood estimator (BNPMLE) $\hat{F}_n(t, z)$ for the bivariate distribution function $F_0(t, z)$ of (T, Z) based on right censored survival data (2.1.2) in which the survival time T is subject to right censoring and the covariate Z is a scalar and is completely observable. This BNPMLE $\hat{F}_n$ is used to facilitate our work in Chapter 3 and provides a starting point of our work in Chapter 4.

In Chapter 3, we derive the empirical likelihood-based confidence interval for conditional survival probabilities with right censored bivariate survival data (2.1.2). We also provide an analytic solution for the empirical likelihood ratio, which is needed for future studies of the asymptotic properties of the empirical likelihood ratio.

In Chapter 4, we extend the weighted empirical likelihood method (Ren, 2001 and 2008a) to the multivariate case and develop estimation methods in a unified form for Cox model (1.2.22) with various types of censored data mentioned in Section 1.3. In particular, we show that the estimator for Cox model (1.2.24) with various types of censored data including those introduced in Section 1.3 is given by the solution of an estimating equation which can be solved using, say, the Newton-Raphson method. As reviewed in Section 1.3.5, currently there has not been any published work on the Cox model with doubly censored data (1.3.6) or with partly interval-censored data (1.3.12)-(1.3.13). Our work here provides solutions to these problems in a unified form.

In Chapter 5, we give some concluding remarks and discuss further research.

CHAPTER 2. BIVARIATE NONPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATOR FOR RIGHT CENSORED DATA

In this chapter, we derive the *bivariate nonparametric maximum likelihood estimator* (BNPMLE) for bivariate d.f. $F_0(t, z)$ of (T, Z), where the survival time T is subject to right censoring and the covariate Z is a scalar and is completely observable.

2.1 Introduction

In survival analysis, we often encounter data in which the survival time T is subject to right censoring and the vector $\boldsymbol{Z} = (Z_1, \cdots, Z_k)$ of covariates such as age, gender, etc., is completely observable. Here, we consider the case where the covariate Z is a scalar, i.e., $k = 1$, and is completely observable. The generalization of our results in this chapter to the case where $\boldsymbol{Z}$ is a vector is straightforward. Suppose that

$$(T_1, Z_1), \cdots, (T_n, Z_n) \overset{\text{iid}}{\sim} F_0(t, z) = P\{T \leq t, Z \leq z\}, \tag{2.1.1}$$

but the actual observed survival data are the bivariate data with the survival time subject to random right censoring as follows:

$$(V_1, \delta_1, Z_1), \cdots, (V_n, \delta_n, Z_n), \tag{2.1.2}$$

where (V_i, δ_i) is right censored data (1.3.2), Z_i is the covariate, and C_i is the right censoring variable with d.f. F_C and density function f_C and is independent of (T_i, Z_i).

In practice, if one wishes to use the nonparametric approach (i.e., without imposing any model assumptions) in the study of the relation between T and Z, a natural thing to do is to estimate the bivariate d.f. $F_0(t,z)$ of (T,Z) based on observed survival data (2.1.2). To our best knowledge, there are no published works on this problem in statistical literature. Another motivation of this work is that the BNPMLE $\hat{F}_n(t,z)$ for $F_0(t,z)$ plays an important role for the weighted empirical likelihood-based estimator for the Cox model (1.2.24) with right censored data (2.1.2), which is studied in Chapter 4 of this dissertation.

In Section 2.2, we derive the BNPMLE $\hat{F}_n(t,z)$ for $F_0(t,z)$ based on right censored bivariate data (2.1.2), and we show that when there is no censoring the BNPMLE coincides with the bivariate empirical d.f. of sample (2.1.1). The proofs are deferred to Section 2.3.

2.2 Bivariate Nonparametric Maximum Likelihood Estimator

To derive the BNPMLE for bivariate d.f. $F_0(t,z)$ of (T,Z) based on right censored data (2.1.2), we let

$$\begin{aligned} &U_1 < \cdots < U_m \text{ be all the distinct observations among } V_1, \cdots, V_n \\ &W_1 < \cdots < W_q \text{ be all the distinct observations among } Z_1, \cdots, Z_n. \end{aligned} \tag{2.2.1}$$

Then, for observed data (2.1.2), the likelihood function is given by

$$\begin{aligned} P\{\text{Observe the given data}\} &= \prod_{k=1}^{n} P\{V=V_k, \delta=\delta_k, Z=Z_k\} \\ &= \prod_{\delta_k=1} P\{T_k=V_k, Z=Z_k, T_k \le C_k\} \prod_{\delta_k=0} P\{C_k=V_k, Z=Z_k, T_k > C_k\} \\ &= \prod_{\delta_k=1} P\{T_k=V_k, Z=Z_k, V_k \le C_k\} \prod_{\delta_k=0} P\{C_k=V_k, Z=Z_k, T_k > V_k\} \\ &\overset{\perp}{=} \prod_{\delta_k=1} P\{T=V_k, Z=Z_k\}\bar{F}_C(V_k) \prod_{\delta_k=0} f_C(V_k) P\{Z=Z_k, T>V_k\} \\ &= \prod_{k=1}^{n} \Big(P\{T=V_k, Z=Z_k\}\bar{F}_C(V_k)\Big)^{\delta_k} \Big(f_C(V_k) P\{T>V_k, Z=Z_k\}\Big)^{1-\delta_k}, \end{aligned} \tag{2.2.2}$$

which is proportional to

$$
\begin{aligned}
&\prod_{k=1}^{n}\Big(P\{T=V_k,Z=Z_k\}\Big)^{\delta_k}\Big(P\{T>V_k,Z=Z_k\}\Big)^{1-\delta_k} \\
&\quad=\prod_{k=1}^{n}\Big(dF_0(V_k,Z_k)\Big)^{\delta_k}\Big(F_0(\infty,dZ_k)-F_0(V_k,dZ_k)\Big)^{1-\delta_k} \\
&\quad=\prod_{i=1}^{m}\prod_{j=1}^{q}\Big(dF_0(U_i,W_j)\Big)^{\gamma_{ij}}\Big(F_0(\infty,dW_j)-F_0(U_i,dW_j)\Big)^{n_{ij}-\gamma_{ij}}, \qquad (2.2.3)
\end{aligned}
$$

where $dF_0(t,z)=P\{T=t,Z=z\}$, $F_0(t,dz)=F_0(t,z)-F_0(t,z-)$, and

$$
n_{ij}=\sum_{k=1}^{n}I\{V_k=U_i,Z_k=W_j\};\quad \gamma_{ij}=\sum_{k=1}^{n}I\{V_k=U_i,\delta_k=1,Z_k=W_j\} \qquad (2.2.4)
$$

for $1\le i\le m,\ 1\le j\le q$. From (2.2.2)-(2.2.3), we see that the likelihood function for d.f. $F_0(t,z)$ of (T,Z) with right censored data (2.1.2) is given by

$$
L(F)=\prod_{i=1}^{m}\prod_{j=1}^{q}\Big(dF(U_i,W_j)\Big)^{\gamma_{ij}}\Big(F(\infty,dW_j)-F(U_i,dW_j)\Big)^{n_{ij}-\gamma_{ij}}, \qquad (2.2.5)
$$

where F is any bivariate d.f., and denoting P_F as the probability under F, we have $dF(t,z)=P_F\{T=t,Z=z\}$ and $F(t,dz)=F(t,z)-F(t,z-)=P_F\{T\le t,Z=z\}$. Note that from (2.2.4) we have the following three consequences:

$$
n=\sum_{i=1}^{m}\sum_{j=1}^{q}n_{ij} \qquad (2.2.6)
$$

$$
n_{1j}+\cdots+n_{mj}\ge 1, \qquad 1\le j\le q \qquad (2.2.7)
$$

$$
n_{ij}=0 \implies \gamma_{ij}=0. \qquad (2.2.8)
$$

This means that those terms with $n_{ij}=0$ in (2.2.5) have no effects to the value of likelihood function $L(F)$. As follows, we deal with this issue as in Ren and Riddlesworth (2011). Let

$$m_j = \max\{i \,|\, n_{ij} > 0\}, \qquad 1 \le j \le q, \tag{2.2.9}$$

then we have from (2.2.8) that $n_{ij} = \gamma_{ij} = 0$ for all $1 \le j \le q$, $m_j < i \le m$; in turn, likelihood function (2.2.5) for F_0 is equivalently written as

$$L(F) = \prod_{j=1}^{q} \prod_{i=1}^{m_j} \Big(dF(U_i, W_j) \Big)^{\gamma_{ij}} \Big(F(\infty, dW_j) - F(U_i, dW_j) \Big)^{n_{ij} - \gamma_{ij}}. \tag{2.2.10}$$

To maximize likelihood function (2.2.10), we restrict all possible candidates to those bivariate d.f.'s that assign all their probability masses to points (U_i, W_j) and line segments $L_j = \{(t, W_j) \in \mathbb{R}^2;\, t > U_m\}$ for $1 \le i \le m$, $1 \le j \le q$. Therefore, likelihood function (2.2.10) becomes

$$L(F) = \prod_{j=1}^{q} \prod_{i=1}^{m_j} (p_{ij})^{\gamma_{ij}} \Big(\sum_{k=i+1}^{m_j+1} p_{kj} \Big)^{n_{ij} - \gamma_{ij}} \;\equiv\; L(\boldsymbol{p}), \tag{2.2.11}$$

where

$$\boldsymbol{p} = \big(p_{11}, \cdots, p_{m_1+1,1}, \cdots, p_{1q}, \cdots, p_{m_q+1,q} \big) \tag{2.2.12}$$

$$F(t, z) = \sum_{i=1}^{m} \sum_{j=1}^{q} q_{ij} I\{U_i \le t, W_j \le z\}, \quad \text{for } t \le U_m,\ z \in \mathbb{R} \tag{2.2.13}$$

satisfy

$$\begin{cases} p_{ij} = q_{ij} = dF(U_i, W_j) = P_F\big\{T = U_i, Z = W_j\big\}, & \text{for } 1 \le j \le q,\ 1 \le i \le m_j \\ q_{m+1,j} = P_F\big\{(T, Z) \in L_j\big\} = P_F\big\{T > U_m, Z = W_j\big\}, & \text{for } 1 \le j \le q \\ p_{m_j+1,j} = P_F\big\{T > U_{m_j}, Z = W_j\big\} = \sum_{i=m_j+1}^{m+1} q_{ij}, & \text{for } 1 \le j \le q \\ \sum_{j=1}^{q} \sum_{i=1}^{m_j+1} p_{ij} = \sum_{j=1}^{q} \sum_{i=1}^{m+1} q_{ij} = 1. \end{cases} \tag{2.2.14}$$

The BNPMLE $\hat{F}_n(t,z)$ for $F_0(t,z)$ is the solution that maximizes the likelihood function $L(F) = L(\boldsymbol{p})$ in (2.2.11) over all functions $F(t,z)$ in (2.2.13) satisfying (2.2.14).

Note that if $m_j < m$, the values of q_{ij}'s for $m_j < i \leq m$ have no effects to the value of likelihood function (2.2.11). Thus, we can only derive the BNPMLE in terms of $\boldsymbol{p}$ (2.2.12) for $L(\boldsymbol{p})$. Let $\hat{\mathbf{p}}$ denote the solution to the following optimization problem:

$$\begin{cases} \max\ L(\boldsymbol{p}) = \prod_{j=1}^{q}\prod_{i=1}^{m_j}(p_{ij})^{\gamma_{ij}}\Big(\sum_{k=i+1}^{m_j+1} p_{kj}\Big)^{n_{ij}-\gamma_{ij}} \\ \text{subject to :}\quad 0 \leq p_{ij} \leq 1,\ \ 1 \leq j \leq q,\ 1 \leq i \leq m_j;\ \ \sum_{j=1}^{q}\sum_{i=1}^{m_j+1} p_{ij} = 1. \end{cases} \tag{2.2.15}$$

The following theorem gives the solution and properties of $\hat{\mathbf{p}}$ with the proof deferred to Section 2.3.

Theorem 2.1. *For any $1 \leq i \leq m$, $1 \leq j \leq q$, we denote*

$$N_{ij} = n_{ij} + \cdots + n_{mj} = \sum_{k=1}^{n} I\{V_k \geq U_i, Z_k = W_j\}. \tag{2.2.16}$$

Then, the solution $\hat{\mathbf{p}}$ of (2.2.15) is unique and satisfies the following:

(i) For any $1 \leq j \leq q,\ 1 \leq i \leq m_j$, we have $\hat{p}_{ij} > 0$ if and only if $\gamma_{ij} > 0$;

(ii) For any $1 \leq j \leq q,\ 1 \leq i \leq m_j$, we have $\sum_{k=i}^{m_j+1} \hat{p}_{kj} > 0$;

(iii) For any $1 \leq j \leq q$, with notation $\prod_{k=1}^{0} c_k \equiv 1$ we have

$$\begin{cases} \hat{p}_{ij} = \Big(\frac{\gamma_{ij}}{N_{ij}}\Big)\Big(\frac{N_{1j}}{n}\Big)\prod_{k=1}^{i-1}\Big(1 - \frac{\gamma_{kj}}{N_{kj}}\Big), \quad \text{for } 1 \leq i \leq m_j \\ \hat{p}_{m_j+1,j} = \frac{N_{1j}}{n} - \sum_{i=1}^{m_j}\hat{p}_{ij}. \end{cases} \tag{2.2.17}$$

It should be noted that Theorem 2.1 shows that the BNPMLE is unique in terms of $\boldsymbol{p}$ (2.2.12), but such uniqueness does not seem obvious in terms of F as given by (2.2.13)-(2.2.14), because if $m_j < m$ and $p_{m_j+1,j} > 0$ for some $1 \leq j \leq q$, it is not obvious how probability mass $p_{m_j+1,j}$ is distributed among $q_{m_j+1,j}, \cdots, q_{mj}, q_{m+1,j}$. In this dissertation, following the treatment in Ren and Riddlesworth (2011) we apply the formula of $\hat{p}_{ij}$'s in (2.2.17) to all $\hat{q}_{ij}$'s generally, then the BNPMLE $\hat{F}_n(t,z)$ for $F_0(t,z)$ is given by

$$\begin{cases} \hat{F}_n(t,z) = \sum_{i=1}^{m}\sum_{j=1}^{q} \hat{q}_{ij} I\{U_i \le t, W_j \le z\}, \quad \text{for } t \le U_m,\ z \in \mathbb{R} \\ \hat{q}_{ij} = \Big(\frac{\gamma_{ij}}{N_{ij}}\Big)\Big(\frac{N_{1j}}{n}\Big)\prod_{k=1}^{i-1}\Big(1-\frac{\gamma_{kj}}{N_{kj}}\Big), \quad \text{for } 1 \le i \le m,\ 1 \le j \le q \\ \hat{q}_{m+1,j} = P_{\hat{F}_n}\big\{T > U_m, Z = W_j\big\} = \frac{N_{1j}}{n} - \sum_{i=1}^{m} \hat{q}_{ij}, \quad \text{for } 1 \le j \le q, \end{cases} \tag{2.2.18}$$

where 0/0 is set as 0 whenever it occurs.

There are two points about (2.2.18) that should be noticed. First, from (2.2.4), (2.2.9), and (2.2.16), we have that

$$\begin{cases} n_{m_j,j} > 0 \implies N_{1j} \ge N_{2j} \ge \cdots \ge N_{m_j,j} > 0, \quad \text{for } 1 \le j \le q \\ n_{ij} = \gamma_{ij} = N_{ij} = 0, \quad \text{for } 1 \le j \le q,\ m_j < i \le m \ \text{ when } m_j < m. \end{cases} \tag{2.2.19}$$

Thus, in line 2 of (2.2.18), we have all $N_{ij} > 0$ for $1 \le i \le m_j$, and that if $m_j < m$, then by line 2 of (2.2.19) we have all $\hat{q}_{ij} = 0$ for $m_j < i \le m$, which means points (U_i, W_j) for $m_j < i \le m$ are not observed among (V_k, Z_k)'s in data (2.1.2), and no probability masses are assigned to these points (U_i, W_j). Second, line 3 of (2.2.18) follows from line 2 of (2.2.17) in two ways: (a) if $m_j = m$, it follows from lines 1-3 of (2.2.14):

$$\hat{q}_{m_j+1,j} = \hat{p}_{m_j+1,j} = \frac{N_{1j}}{n} - \sum_{i=1}^{m_j} \hat{p}_{ij} = \frac{N_{1j}}{n} - \sum_{i=1}^{m_j} \hat{q}_{ij};$$

(b) if $m_j < m$, it follows from $\hat{p}_{m_j+1,j} = \sum_{i=m_j+1}^{m+1} \hat{q}_{ij} = \hat{q}_{m+1,j}$ and

$$\hat{p}_{m_j+1,j} = \frac{N_{1j}}{n} - \sum_{i=1}^{m_j} \hat{p}_{ij} = \frac{N_{1j}}{n} - \sum_{i=1}^{m_j} \hat{q}_{ij},$$

which are due to lines 1 and 3 in (2.2.14) and the fact that $\hat{q}_{ij} = 0$ for all $m_j < i \le m$.

The following corollary shows that the BNPMLE $\hat{F}_n(t,z)$ in (2.2.18) coincides with the bivariate empirical d.f. when there is no censoring in (2.1.2).

Corollary 2.1. *When there is no censoring in data (2.1.2), BNPMLE $\hat{F}_n(t,z)$ in (2.2.18) coincides with the bivariate empirical d.f. of sample (2.1.1).*

Remark 2.1: From (2.2.18), the probability under $\hat{F}_n$ at any given point W_α for some $1 \le \alpha \le q$ is given as follows:

$$\begin{aligned} P_{\hat{F}_n}\{Z = W_\alpha\} &= P_{\hat{F}_n}\{T \le U_m,\ Z = W_\alpha\} + P_{\hat{F}_n}\{T > U_m,\ Z = W_\alpha\} \\ &= \sum_{i=1}^{m}\sum_{j=1}^{q} \hat{q}_{ij} I\{W_j = W_\alpha\} + \hat{q}_{m+1,\alpha} \\ &= \sum_{i=1}^{m} \hat{q}_{i\alpha} + \frac{N_{1j}}{n} - \sum_{i=1}^{m} \hat{q}_{i\alpha} = \frac{N_{1j}}{n}. \end{aligned} \tag{2.2.20}$$

2.3 Proofs

Proof of Theorem 2.1 (i): "$\Leftarrow$": Clearly, if $\gamma_{\alpha\zeta} > 0$ for some $1 \le \zeta \le q,\ 1 \le \alpha \le m_\zeta$, then any solution $\boldsymbol{p}$ that maximizes $L(\boldsymbol{p})$ in (2.2.15) satisfies $p_{\alpha\zeta} > 0$.

"$\Rightarrow$": Assume $\gamma_{\alpha\zeta} = 0$ for some $1 \le \zeta \le q,\ 1 \le \alpha \le m_\zeta$. Then, likelihood function (2.2.15) is equivalently written as

$$\begin{aligned} L(\boldsymbol{p}) = &\Big(\prod_{\substack{j=1 \\ j\ne\zeta}}^{q}\prod_{i=1}^{m_j} (p_{ij})^{\gamma_{ij}} \Big(\sum_{k=i+1}^{m_j+1} p_{kj}\Big)^{n_{ij}-\gamma_{ij}}\Big)\Big(p_{\alpha+1,\zeta} + \cdots + p_{m_\zeta+1,\zeta}\Big)^{n_{\alpha\zeta}} \\ &\times \Big(\prod_{i=1}^{\alpha-1} (p_{i\zeta})^{\gamma_{i\zeta}}\Big(p_{i+1,\zeta} + \cdots + p_{\alpha\zeta} + \cdots + p_{m_\zeta+1,\zeta}\Big)^{n_{i\zeta}-\gamma_{i\zeta}}\Big) \\ &\times \Big(\prod_{i=\alpha+1}^{m_\zeta} (p_{i\zeta})^{\gamma_{i\zeta}}\Big(p_{i+1,\zeta} + \cdots + p_{m_\zeta+1,\zeta}\Big)^{n_{i\zeta}-\gamma_{i\zeta}}\Big), \end{aligned} \tag{2.3.1}$$

where from (2.2.9) and $\alpha \le m_\zeta$ we know $n_{\alpha\zeta} > 0$. Assume $\hat{p}_{\alpha\zeta} > 0$, where $\hat{\boldsymbol{p}}$ is a solution to (2.2.15), and let $\tilde{\boldsymbol{p}}$ be a vector with components $\tilde{p}_{ij}$, $1 \le j \le q,\ 1 \le i \le m_j$, that satisfy

$$\tilde{p}_{\alpha\zeta} = 0; \qquad \tilde{p}_{m_\zeta+1,\zeta} = \hat{p}_{m_\zeta+1,\zeta} + \hat{p}_{\alpha\zeta}; \qquad \tilde{p}_{ij} = \hat{p}_{ij}, \ \text{ otherwise}, \tag{2.3.2}$$

which implies

$$\begin{cases} 0 \le \tilde{p}_{ij} \le 1, \qquad 1 \le j \le q,\ 1 \le i \le m_j \\ \sum_{j=1}^{q} \sum_{i=1}^{m_j+1} \tilde{p}_{ij} = \sum_{j=1}^{q} \sum_{i=1}^{m_j+1} \hat{p}_{ij} = 1 \\ \hat{p}_{m_\zeta+1,\zeta} < \tilde{p}_{m_\zeta+1,\zeta}, \qquad \hat{p}_{m_\zeta+1,\zeta} + \hat{p}_{\alpha\zeta} = \tilde{p}_{m_\zeta+1,\zeta} + \tilde{p}_{\alpha\zeta}. \end{cases} \tag{2.3.3}$$

Therefore, from (2.3.1)-(2.3.3), we have

$$\begin{aligned} L(\hat{\boldsymbol{p}}) = & \Bigg(\prod_{\substack{j=1 \\ j\neq\zeta}}^{q} \prod_{i=1}^{m_j} (\hat{p}_{ij})^{\gamma_{ij}} \Big(\sum_{k=i+1}^{m_j+1} \hat{p}_{kj} \Big)^{n_{ij}-\gamma_{ij}} \Bigg) \Big(\hat{p}_{\alpha+1,\zeta} + \cdots + \hat{p}_{m_\zeta+1,\zeta} \Big)^{n_{\alpha\zeta}} \\ & \times \Bigg(\prod_{i=1}^{\alpha-1} (\hat{p}_{i\zeta})^{\gamma_{i\zeta}} \Big(\hat{p}_{i+1,\zeta} + \cdots + \hat{p}_{\alpha\zeta} + \cdots + \hat{p}_{m_\zeta+1,\zeta} \Big)^{n_{i\zeta}-\gamma_{i\zeta}} \Bigg) \\ & \times \Bigg(\prod_{i=\alpha+1}^{m_\zeta} (\hat{p}_{i\zeta})^{\gamma_{i\zeta}} \Big(\hat{p}_{i+1,\zeta} + \cdots + \hat{p}_{m_\zeta+1,\zeta} \Big)^{n_{i\zeta}-\gamma_{i\zeta}} \Bigg) \\ < & \Bigg(\prod_{\substack{j=1 \\ j\neq\zeta}}^{q} \prod_{i=1}^{m_j} (\tilde{p}_{ij})^{\gamma_{ij}} \Big(\sum_{k=i+1}^{m_j+1} \tilde{p}_{kj} \Big)^{n_{ij}-\gamma_{ij}} \Bigg) \Big(\tilde{p}_{\alpha+1,\zeta} + \cdots + \tilde{p}_{m_\zeta+1,\zeta} \Big)^{n_{\alpha\zeta}} \\ & \times \Bigg(\prod_{i=1}^{\alpha-1} (\tilde{p}_{i\zeta})^{\gamma_{i\zeta}} \Big(\tilde{p}_{i+1,\zeta} + \cdots + \tilde{p}_{\alpha\zeta} + \cdots + \tilde{p}_{m_\zeta+1,\zeta} \Big)^{n_{i\zeta}-\gamma_{i\zeta}} \Bigg) \\ & \times \Bigg(\prod_{i=\alpha+1}^{m_\zeta} (\tilde{p}_{i\zeta})^{\gamma_{i\zeta}} \Big(\tilde{p}_{i+1,\zeta} + \cdots + \tilde{p}_{m_\zeta+1,\zeta} \Big)^{n_{i\zeta}-\gamma_{i\zeta}} \Bigg) \ \equiv \ L(\tilde{\boldsymbol{p}}), \end{aligned}$$

a contradiction. $\square$

Proof of Theorem 2.1 (ii): Note that for any solution $\boldsymbol{p}$ that maximizes $L(\boldsymbol{p})$ in (2.2.15), we have

$$L(\boldsymbol{p}) > 0,$$

which implies the following in the product of $L(\boldsymbol{p})$ in (2.2.15):

$$\begin{aligned} 0 &< (p_{m_j,j})^{\gamma_{m_j,j}} (p_{m_j+1,j})^{n_{m_j,j}-\gamma_{m_j,j}} \\ &= \begin{cases} (p_{m_j+1,j})^{n_{m_j,j}} & \text{if} \quad \gamma_{m_j,j} = 0 \\ (p_{m_j,j})^{\gamma_{m_j,j}} (p_{m_j+1,j})^{n_{m_j,j}-\gamma_{m_j,j}} & \text{if} \quad \gamma_{m_j,j} > 0. \end{cases} \end{aligned}$$

Therefore, since $n_{m_j,j} > 0$ in (2.2.19), we have $p_{m_j+1,j} > 0$ when $\gamma_{m_j,j} = 0$ and $p_{m_j,j} > 0$ when $\gamma_{m_j,j} > 0$, which imply $p_{m_j,j} + p_{m_j+1,j} > 0$. $\square$

Before proving Theorem 2.1 (iii), we establish the following lemmas, while the proofs are given at the end of this section.

Lemma 2.1. *Let*

$$J_1 \equiv \left\{(i,j) \,|\, \gamma_{ij} = 0,\ N_{ij} - \gamma_{ij} > 0,\ 1 \le j \le q,\ 1 \le i \le m_j\right\} \tag{2.3.4}$$

$$G_1(\boldsymbol{a}_1) \equiv \prod\prod_{(i,j)\in J_1} (1-a_{ij})^{N_{ij}}, \qquad \boldsymbol{a}_1 = \{a_{ij} \,|\, (i,j) \in J_1\}. \tag{2.3.5}$$

Then, the solution to

$$\begin{cases} \max\ G_1(\boldsymbol{a}_1) \\ \text{subject to}: \quad 0 \le a_{ij} \le 1, \quad (i,j) \in J_1 \end{cases} \tag{2.3.6}$$

is uniquely given by

$$\hat{\boldsymbol{a}}_1 = \mathbf{0}. \tag{2.3.7}$$

Lemma 2.2. *Let*

$$J_2 \equiv \left\{(i,j)\,|\,\gamma_{ij} > 0,\ N_{ij} - \gamma_{ij} = 0,\ 1 \le j \le q,\ 1 \le i \le m_j\right\} \tag{2.3.8}$$

$$G_2(\boldsymbol{a}_2) \equiv \prod_{(i,j)\in J_2}\prod (a_{ij})^{\gamma_{ij}}, \qquad \boldsymbol{a}_2 = \{a_{ij}\,|\,(i,j) \in J_2\}. \tag{2.3.9}$$

Then, the solution to

$$\begin{cases} \max\ G_2(\boldsymbol{a}_2) \\ \text{subject to}: \quad 0 \le a_{ij} \le 1, \quad (i,j) \in J_2 \end{cases} \tag{2.3.10}$$

is uniquely given by

$$\hat{\boldsymbol{a}}_2 = \mathbf{1}. \tag{2.3.11}$$

Lemma 2.3. *Let*

$$J_3 \equiv \left\{(i,j)\,|\,\gamma_{ij} > 0,\ N_{ij} - \gamma_{ij} > 0,\ 1 \le j \le q,\ 1 \le i \le m_j\right\} \tag{2.3.12}$$

$$G_3(\boldsymbol{a}_3) \equiv \prod_{(i,j)\in J_3}\prod (a_{ij})^{\gamma_{ij}}(1 - a_{ij})^{N_{ij}-\gamma_{ij}}, \qquad \boldsymbol{a}_3 = \{a_{ij}\,|\,(i,j) \in J_3\}. \tag{2.3.13}$$

Then, the solution to

$$\begin{cases} \max\ G_3(\boldsymbol{a}_3) \\ \text{subject to}: \quad 0 \le a_{ij} \le 1, \quad (i,j) \in J_3. \end{cases} \tag{2.3.14}$$

is uniquely given by $\hat{\boldsymbol{a}}_3 = \left\{\hat{a}_{ij}\,|\,(i,j) \in J_3\right\}$*, where*

$$\hat{a}_{ij} = \frac{\gamma_{ij}}{N_{ij}}, \qquad (i,j) \in J_3. \tag{2.3.15}$$

Lemma 2.4. *Let*

$$J \equiv \big\{(i,j) \,|\, 1 \le j \le q,\ 1 \le i \le m_j\big\} \tag{2.3.16}$$

$$G_4(\boldsymbol{a}) \equiv \prod_{(i,j)\in J}\prod (a_{ij})^{\gamma_{ij}}(1-a_{ij})^{N_{ij}-\gamma_{ij}}, \qquad \boldsymbol{a} = \{a_{ij} \,|\, (i,j) \in J\}. \tag{2.3.17}$$

Then, the solution to

$$\begin{cases} \max\ G_4(\boldsymbol{a}) \\ \text{subject to :} \quad 0 \le a_{ij} \le 1, \quad (i,j) \in J. \end{cases} \tag{2.3.18}$$

is given by $\hat{\boldsymbol{a}} = \big\{\hat{a}_{ij} \,|\, (i,j) \in J\big\}$, *where*

$$\hat{a}_{ij} = \frac{\gamma_{ij}}{N_{ij}}, \qquad (i,j) \in J. \tag{2.3.19}$$

Lemma 2.5. *Let*

$$G_5(\boldsymbol{b}) \equiv \prod_{j=1}^{q} (b_{1j})^{N_{1j}}, \qquad \boldsymbol{b} = (b_{11}, \cdots, b_{1q}). \tag{2.3.20}$$

Then, the solution to

$$\begin{cases} \max\ G_5(\boldsymbol{b}) \\ \text{subject to :} \quad 0 \le b_{1j} \le 1, \quad 1 \le j \le q; \quad \sum_{j=1}^{q} b_{1j} = 1, \end{cases} \tag{2.3.21}$$

is uniquely given by $\hat{\boldsymbol{b}} = (\hat{b}_{11}, \cdots, \hat{b}_{1q})$, *where*

$$\hat{b}_{1j} = \frac{N_{1j}}{n}, \qquad 1 \le j \le q. \tag{2.3.22}$$

Proof of Theorem 2.1 (iii): To find the solution to (2.2.15), consider the following substitutions:

$$a_{ij} = \frac{p_{ij}}{b_{ij}} \qquad \text{and} \qquad b_{ij} = \sum_{k=i}^{m_j+1} p_{kj}, \qquad \text{for } \ 1 \leq j \leq q,\ 1 \leq i \leq m_j, \tag{2.3.23}$$

which imply

$$\begin{cases} b_{m_j+1,j} = p_{m_j\,+\,1,\,j} \\ b_{i+1,j} = \sum_{k=i+1}^{m_j+1} p_{kj} = b_{ij} - p_{ij} \\ 1 - a_{ij} = \dfrac{b_{i+1,j}}{b_{ij}}. \end{cases} \tag{2.3.24}$$

Therefore, $L(\boldsymbol{p})$ in (2.2.15) can be equivalently written as

$$\begin{aligned} L(\boldsymbol{p}) &= \prod_{j=1}^{q}\prod_{i=1}^{m_j} (p_{ij})^{\gamma_{ij}} (b_{i+1,j})^{n_{ij}-\gamma_{ij}} \ = \ \prod_{j=1}^{q}\prod_{i=1}^{m_j} (p_{ij})^{\gamma_{ij}} (b_{ij} - p_{ij})^{n_{ij}-\gamma_{ij}} \\ &= \prod_{j=1}^{q}\prod_{i=1}^{m_j} (p_{ij})^{\gamma_{ij}} \Big(\frac{p_{ij}}{a_{ij}} - p_{ij}\Big)^{n_{ij}-\gamma_{ij}} \ = \ \prod_{j=1}^{q}\prod_{i=1}^{m_j} (p_{ij})^{n_{ij}} \prod_{j=1}^{q}\prod_{i=1}^{m_j} \Big(\frac{1-a_{ij}}{a_{ij}}\Big)^{n_{ij}-\gamma_{ij}} \\ &= \Big(\prod_{j=1}^{q}\prod_{i=1}^{m_j} (p_{ij})^{n_{ij}}\Big) \frac{\prod_{j=1}^{q}\prod_{i=1}^{m_j}\ (a_{ij})^{\gamma_{ij}} (1-a_{ij})^{N-\gamma_{ij}-(n_{1j}+\cdots+n_{i-1,j})}}{\prod_{j=1}^{q}\prod_{i=1}^{m_j}\ (a_{ij})^{n_{ij}} (1-a_{ij})^{N-(n_{1j}+\cdots+n_{ij})}} \\ &= \Big(\prod_{j=1}^{q}\prod_{i=1}^{m_j} (p_{ij})^{n_{ij}}\Big) \frac{\prod_{j=1}^{q}\prod_{i=1}^{m_j}\ (a_{ij})^{\gamma_{ij}} (1-a_{ij})^{N-\gamma_{ij}-(N_{1j}-N_{ij})}}{\prod_{j=1}^{q}\prod_{i=1}^{m_j}\ (a_{ij})^{n_{ij}} (1-a_{ij})^{N-(N_{1j}-N_{i+1,j})}} \\ &= \Big(\prod_{j=1}^{q}\prod_{i=1}^{m_j} (b_{ij})^{n_{ij}}\Big) \frac{\prod_{j=1}^{q}\prod_{i=1}^{m_j}\ (a_{ij})^{\gamma_{ij}} (1-a_{ij})^{N-\gamma_{ij}-(N_{1j}-N_{ij})}}{\prod_{j=1}^{q}\prod_{i=1}^{m_j}\ (1-a_{ij})^{N-(N_{1j}-N_{i+1,j})}}, \end{aligned} \tag{2.3.25}$$

where $N \equiv m \cdot q$ and N_{ij} is given by (2.2.16). From (2.3.24) the denominator in (2.3.25) can be written as

$$\begin{aligned} \prod_{j=1}^{q}\prod_{i=1}^{m_j} (1-a_{ij})^{N-(N_{1j}-N_{i+1,j})} \ &= \ \prod_{j=1}^{q}\prod_{i=1}^{m_j} \Big(\frac{b_{i+1,j}}{b_{ij}}\Big)^{N-(n_{1j}+\cdots+n_{ij})} \\ &= \prod_{j=1}^{q} \Big(\frac{(b_{2j})^{N-n_{1j}}}{(b_{1j})^{N-n_{1j}}}\ \frac{(b_{3j})^{N-n_{1j}-n_{2j}}}{(b_{2j})^{N-n_{1j}-n_{2j}}} \cdot \cdots \cdot \frac{(b_{m_j+1,j})^{N-(n_{1j}+\cdots+n_{m_j,j})}}{(b_{m_j,j})^{N-(n_{1j}+\cdots+n_{m_j,j})}}\Big) \end{aligned}$$

$$
\begin{aligned}
&= \prod_{j=1}^{q} \frac{1}{(b_{1j})^N} \Big((b_{1j})^{n_{1j}} (b_{2j})^{n_{2j}} \cdots (b_{m_j,j})^{n_{m_j,j}} \Big) (b_{m_j+1,j})^{N-(N_{1j}-N_{m_j+1,j})} \\
&= \Big(\prod_{j=1}^{q} \prod_{i=1}^{m_j} (b_{ij})^{n_{ij}} \Big) \prod_{j=1}^{q} \frac{(b_{m_j+1,j})^{N-(N_{1j}-N_{m_j+1,j})}}{(b_{1j})^N} \\
&= \Big(\prod_{j=1}^{q} \prod_{i=1}^{m_j} (b_{ij})^{n_{ij}} \Big) \prod_{j=1}^{q} \frac{(b_{m_j+1,j})^{N-N_{1j}}}{(b_{1j})^N},
\end{aligned}
\tag{2.3.26}
$$

where the last equality is true because from (2.2.16) and (2.2.19), we have $N_{m_j+1,j} = 0$ for all $1 \leq j \leq q$. Since $N - N_{1j}$ is independent of i, we have, from (2.3.24),

$$
\begin{aligned}
\prod_{j=1}^{q} \prod_{i=1}^{m_j} (1-a_{ij})^{N-N_{1j}} &= \prod_{j=1}^{q} \Big(\prod_{i=1}^{m_j} (1-a_{ij}) \Big)^{N-N_{1j}} \\
= \prod_{j=1}^{q} \Big(\prod_{i=1}^{m_j} \frac{b_{i+1,j}}{b_{ij}} \Big)^{N-N_{1j}} &= \prod_{j=1}^{q} \Big(\frac{b_{2j}}{b_{1j}} \; \frac{b_{3j}}{b_{2j}} \cdot \cdots \frac{b_{m_j+1,j}}{b_{m_j,j}} \Big)^{N-N_{1j}} \\
= \prod_{j=1}^{q} \Big(\frac{b_{m_j+1,j}}{b_{1j}} \Big)^{N-N_{1j}} &= \frac{\prod_{j=1}^{q} (b_{m_j+1,j})^{N-N_{1j}}}{\prod_{j=1}^{q} (b_{1j})^{N-N_{1j}}},
\end{aligned}
$$

which implies

$$
\prod_{j=1}^{q} (b_{m_j+1,j})^{N-N_{1j}} = \Big(\prod_{j=1}^{q} (b_{1j})^{N-N_{1j}} \Big) \prod_{j=1}^{q} \prod_{i=1}^{m_j} (1-a_{ij})^{N-N_{1j}}. \tag{2.3.27}
$$

Plugging (2.3.27) into (2.3.26), we obtain

$$
\prod_{j=1}^{q} \prod_{i=1}^{m_j} (1-a_{ij})^{N-(N_{1j}-N_{i+1,j})} = \Big(\prod_{j=1}^{q} \prod_{i=1}^{m_j} (b_{ij})^{n_{ij}} \Big) \Big(\frac{\prod_{j=1}^{q} \prod_{i=1}^{m_j} (1-a_{ij})^{N-N_{1j}}}{\prod_{j=1}^{q} (b_{1j})^{N_{1j}}} \Big),
$$

in turn, $L(\boldsymbol{p})$ in (2.3.25) is equivalently written as

$$
\begin{aligned}
L(\boldsymbol{p}) &= \Big(\prod_{j=1}^{q} (b_{1j})^{N_{1j}} \Big) \prod_{j=1}^{q} \prod_{i=1}^{m_j} (a_{ij})^{\gamma_{ij}} (1-a_{ij})^{N_{ij}-\gamma_{ij}} \\
&= \Big(\prod_{j=1}^{q} (b_{1j})^{N_{1j}} \Big) \prod\prod_{(i,j)\in J} (a_{ij})^{\gamma_{ij}} (1-a_{ij})^{N_{ij}-\gamma_{ij}} \equiv G(\boldsymbol{a}, \boldsymbol{b}),
\end{aligned}
\tag{2.3.28}
$$

where J is given by (2.3.16) and

$$\boldsymbol{a} = \big\{a_{ij} \,|\, (i,j) \in J\big\}; \qquad \boldsymbol{b} = (b_{11}, \dots, b_{1q}), \tag{2.3.29}$$

with a_{ij} and b_{1j} given by (2.3.23). Therefore, optimization problem (2.2.15) is equivalent to the following optimization problem:

$$\begin{cases} \max G(\boldsymbol{a}, \boldsymbol{b}) \\ \text{subject to :} \quad 0 \le a_{ij} \le 1,\ 0 \le b_{1j} \le 1,\ (i,j) \in J; \quad \sum_{j=1}^{q} b_{1j} = 1, \end{cases} \tag{2.3.30}$$

where we know a solution to (2.3.30) exists because $G(\boldsymbol{a}, \boldsymbol{b})$ in (2.3.28) is a polynomial, thus is continuous in $(\boldsymbol{a}, \boldsymbol{b})$, and the constraint set

$$\Big\{(\boldsymbol{a}, \boldsymbol{b}) \,\Big|\, 0 \le a_{ij} \le 1,\ 0 \le b_{1j} \le 1,\ (i,j) \in J;\ \sum_{j=1}^{q} b_{1j} = 1\Big\}$$

is compact. Since the constraint set can be written as

$$\begin{aligned} &\Big\{(\boldsymbol{a}, \boldsymbol{b}) \,\Big|\, 0 \le a_{ij} \le 1,\ 0 \le b_{1j} \le 1,\ (i,j) \in J;\ \sum_{j=1}^{q} b_{1j} = 1\Big\} \\ &= \Big\{\boldsymbol{a} \,\big|\, 0 \le a_{ij} \le 1,\ (i,j) \in J\Big\} \cup \Big\{\boldsymbol{b} \,\Big|\, 0 \le b_{1j} \le 1,\ 1 \le j \le q;\ \sum_{j=1}^{q} b_{1j} = 1\Big\} \end{aligned}$$

and $G(\boldsymbol{a}, \boldsymbol{b})$ in (2.3.28) can be written as

$$G(\boldsymbol{a}, \boldsymbol{b}) = G_4(\boldsymbol{a})\, G_5(\boldsymbol{b}), \tag{2.3.31}$$

where $G_4(\boldsymbol{a})$ and $G_5(\boldsymbol{b})$ are given by (2.3.17) and (2.3.20), respectively, we know from Lemmas 2.4 and 2.5 that the solution for (2.3.30) is uniquely given by $(\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}})$, where $\hat{\boldsymbol{a}}$ is the unique solution to optimization problem (2.3.18) and is given by (2.3.19) and $\hat{\boldsymbol{b}}$ is the unique solution to optimization problem (2.3.21) and is given by (2.3.22).

To obtain an expression for p_{ij} in terms of a_{ij} and b_{1j}, note that from (2.3.23), we have

$$
\begin{aligned}
p_{1j} &= a_{1j}b_{1j}, \\
p_{2j} &= a_{2j}b_{2j} = a_{2j}(b_{1j} - p_{1j}) = a_{2j}b_{1j}(1 - a_{1j}), \\
p_{3j} &= a_{3j}b_{3j} = a_{3j}(b_{1j} - p_{1j} - p_{2j}) = a_{3j}(b_{1j} - a_{1j}b_{1j} - a_{2j}b_{1j}(1 - a_{1j})) \\
&= a_{3j}b_{1j}(1 - a_{1j})(1 - a_{2j}) = a_{3j}b_{1j}\prod_{k=1}^{2}(1 - a_{kj}),
\end{aligned}
$$

and if we continue this we get the following general expression for p_{ij}:

$$
p_{ij} = a_{ij}b_{1j}\prod_{k=1}^{i-1}(1 - a_{kj}), \qquad 1 \le j \le q, \;\; 1 \le i \le m_j, \tag{2.3.32}
$$

where $\prod_{k=1}^{0} c_k \equiv 1$. Also, from (2.3.23), we have

$$
p_{m_j+1,j} = b_{1j} - \sum_{i=1}^{m_j} p_{ij}, \qquad 1 \le j \le q. \tag{2.3.33}
$$

Therefore, the solution to (2.2.15) is uniquely given by

$$
\begin{cases}
\hat{p}_{ij} = \hat{a}_{ij}\hat{b}_{1j}\prod_{k=1}^{i-1}(1 - \hat{a}_{kj}), & 1 \le j \le q, \;\; 1 \le i \le m_j \\
\hat{p}_{m_j+1,j} = \hat{b}_{1j} - \sum_{i=1}^{m_j} \hat{p}_{ij}, & 1 \le j \le q,
\end{cases} \tag{2.3.34}
$$

where $(\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}})$ is the unique solution to (2.3.30). The proof follows from (2.3.34) with $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$ given by (2.3.19) and (2.3.22), respectively. $\square$

Proof of Corollary 2.1: As follows, we find $\hat{q}_{ij}$ in (2.2.18) when there is no censoring in data (2.1.2). Without loss of generality, assume $T_1 < T_2 < \cdots < T_n$. Since there is no censoring in data (2.1.2), the observed data are

$$(T_i, 1, Z_i), \qquad i = 1, \ldots, n, \tag{2.3.35}$$

and we have $U_i = T_i$ and $W_j = Z_k$ for some $1 \le k \le n$ and all $1 \le i \le m$, $1 \le j \le q$ with $m = n$ and $q \le n$. Note that from (2.2.19) we know

$$\hat{q}_{ij} = 0, \qquad 1 \le j \le q,\ m_j < i \le m. \tag{2.3.36}$$

Suppose $1 \le j \le q$, $1 \le i \le m_j$. Since there is no censoring in data (2.1.2), we have from $T_1 < T_2 < \cdots < T_n$ and (2.2.4)

$$\gamma_{ij} = n_{ij} = 0 \ \text{ or } \ 1 \qquad \text{for all } \ 1 \le j \le q,\ 1 \le i \le m_j. \tag{2.3.37}$$

If $\gamma_{ij} = 0$, we have in (2.2.18)

$$\hat{q}_{ij} = 0. \tag{2.3.38}$$

Consider the case $\gamma_{ij} = n_{ij} = 1$, which implies that $\hat{q}_{ij}$ in (2.2.18) becomes

$$\hat{q}_{ij} = \Big(\frac{1}{N_{ij}}\Big)\Big(\frac{N_{1j}}{n}\Big)\prod_{k=1}^{i-1}\Big(1 - \frac{\gamma_{kj}}{N_{kj}}\Big). \tag{2.3.39}$$

Since $N_{1j} \ge N_{ij}$, there are two possible cases:

Case 1: $N_{1j} = N_{ij}$

Case 2: $N_{1j} > N_{ij}$.

As follows, we find $\hat{q}_{ij}$ given by (2.3.39) for these two cases, respectively.

Case 1: Since $N_{1j} = N_{ij}$, $\hat{q}_{ij}$ in (2.3.39) becomes

$$\hat{q}_{ij} = \frac{1}{n} \prod_{k=1}^{i-1} \Big(1 - \frac{\gamma_{kj}}{N_{kj}}\Big). \tag{2.3.40}$$

If $i = 1$, we have

$$\hat{q}_{ij} = \frac{1}{n} \prod_{k=1}^{0} \Big(1 - \frac{\gamma_{kj}}{N_{kj}}\Big) = \frac{1}{n}(1) = \frac{1}{n}. \tag{2.3.41}$$

Suppose $i > 1$. From $N_{1j} = N_{ij}$ and (2.2.16), we have

$$n_{1j} + \cdots + n_{i-1,j} + n_{ij} + \cdots + n_{nj} \ = \ n_{ij} + \cdots + n_{nj},$$

which implies $n_{1j} = \cdots = n_{i-1,j} = 0$. Therefore, $\gamma_{1j} = \cdots = \gamma_{i-1,j} = 0$ from (2.3.37), and $\hat{q}_{ij}$ in (2.3.40) becomes

$$\hat{q}_{ij} = \frac{1}{n} \prod_{k=1}^{i-1} \Big(1 - \frac{0}{N_{kj}}\Big) = \frac{1}{n}(1) = \frac{1}{n}. \tag{2.3.42}$$

Case 2: Since $N_{1j} > N_{ij}$, we have from (2.2.16)

$$n_{1j} + \cdots + n_{i-1,j} + n_{ij} + \cdots + n_{nj} \ > \ n_{ij} + \cdots + n_{nj},$$

which implies that $i \neq 1$ and there is at least one $1 \leq k \leq i-1$ such that $n_{kj} = 1$. Therefore, there exist $1 \leq i_1 < \cdots < i_\ell \leq i-1$ such that

$$\begin{cases} n_{i_1 j} = \cdots = n_{i_\ell j} = 1 \\ n_{kj} = 0 \quad \text{for all} \quad 1 \leq k \leq i-1, \quad k \neq i_1, \ldots, i_\ell. \end{cases} \tag{2.3.43}$$

Thus, from (2.3.37), we see that

$$\begin{cases} \gamma_{i_1 j} = \cdots = \gamma_{i_\ell j} = 1 \\ \gamma_{kj} = 0 \quad \text{for} \;\; \text{all} \;\; 1 \le k \le i-1, \quad k \ne i_1, \dots, i_\ell, \end{cases} \tag{2.3.44}$$

and $\hat{q}_{ij}$ in (2.3.39) becomes

$$\begin{aligned} \hat{q}_{ij} &= \Big(\frac{1}{N_{ij}}\Big)\Big(\frac{N_{1j}}{n}\Big)\prod_{k=1}^{i-1}\Big(1-\frac{\gamma_{kj}}{N_{kj}}\Big) \\ &= \Big(\frac{1}{N_{ij}}\Big)\Big(\frac{N_{1j}}{n}\Big)\Big(1-\frac{1}{N_{i_1 j}}\Big)\Big(1-\frac{1}{N_{i_2 j}}\Big)\cdots\Big(1-\frac{1}{N_{i_\ell j}}\Big) \\ &= \Big(\frac{1}{N_{ij}}\Big)\Big(\frac{N_{1j}}{n}\Big)\Big(\frac{N_{i_1 j}-1}{N_{i_1 j}}\Big)\Big(\frac{N_{i_2 j}-1}{N_{i_2 j}}\Big)\cdots\Big(\frac{N_{i_\ell j}-1}{N_{i_\ell j}}\Big) \\ &= \Big(\frac{1}{n}\Big)\Big(\frac{N_{1j}}{N_{i_1 j}}\Big)\Big(\frac{N_{i_1 j}-1}{N_{i_2 j}}\Big)\Big(\frac{N_{i_2 j}-1}{N_{i_3 j}}\Big)\cdots\Big(\frac{N_{i_\ell j}-1}{N_{ij}}\Big), \end{aligned} \tag{2.3.45}$$

where the last equality is a rearrangement of the terms. Note that from (2.3.43), we have

$$\begin{aligned} &N_{i_1 j} = n_{i_1 j} + \cdots + n_{i-1,j} + n_{ij} + \cdots + n_{nj} = \underbrace{1+\cdots+1}_{\ell \text{ times}} + N_{ij} = \ell + N_{ij} = N_{1j} \\ &N_{i_2 j} = n_{i_2 j} + \cdots + n_{i-1,j} + n_{ij} + \cdots + n_{nj} = \underbrace{1+\cdots+1}_{\ell-1 \text{ times}} + N_{ij} = \ell - 1 + N_{ij} = N_{i_1 j} - 1 \\ &\qquad \vdots \\ &N_{i_\ell j} = n_{i_\ell j} + \cdots + n_{i-1,j} + n_{ij} + \cdots + n_{nj} = 1 + N_{ij} = N_{i_{\ell-1},j} - 1 \\ &N_{ij} = n_{ij} + \cdots + n_{nj} = N_{i_\ell j} - 1, \end{aligned}$$

which implies from (2.3.45)

$$\hat{q}_{ij} = \frac{1}{n}. \tag{2.3.46}$$

This ends the argument for Case 2.

From (2.3.36), (2.3.38), (2.3.41)-(2.3.42) and (2.3.46), we have

$$\hat{q}_{ij} = \begin{cases} 0, & \text{if } \ 1 \le j \le q, \ m_j < i \le m \\ 0, & \text{if } \ \gamma_{ij} = 0, \ 1 \le j \le q, \ 1 \le i \le m_j \\ \dfrac{1}{n}, & \text{if } \ \gamma_{ij} = 1, \ 1 \le j \le q, \ 1 \le i \le m_j. \end{cases}$$

$$= \begin{cases} 0, & \text{if } \ \gamma_{ij} = 0, \ 1 \le j \le q, \ 1 < i \le n \\ \dfrac{1}{n}, & \text{if } \ \gamma_{ij} = 1, \ 1 \le j \le q, \ 1 < i \le n, \end{cases} \tag{2.3.47}$$

where the last equality is true because $m = n$ and $\gamma_{ij} = 0$ for $m_j < i \le m$ from (2.2.19). Therefore, when there is no censoring in data (2.1.2), $\hat{F}_n(t,z)$ in (2.2.18) becomes

$$\hat{F}_n(t,z) = \frac{1}{n} \sum_{\substack{j=1 \\ \gamma_{ij}=1}}^{q} \sum_{i=1}^{m_j} I\{U_i \le t, W_j \le z\} = \frac{1}{n} \sum_{\substack{j=1 \\ \gamma_{ij}=1}}^{q} \sum_{i=1}^{n} I\{U_i \le t, W_j \le z\} \tag{2.3.48}$$

$$t \le U_m, \ z \in \mathbb{R}.$$

But from (2.2.4) and (2.3.35) and since there is no censoring in data (2.1.2), we have

$$\Big\{(i,j) \,\big|\, \gamma_{ij} = 1\Big\} = \Big\{(i,j) \,\Big|\, \sum_{k=1}^{n} I\{T_k = U_i, \ Z_k = W_j\} = 1\Big\}$$

$$= \Big\{(i,j) \,\big|\, (U_i, W_j) = (T_{k_0}, Z_{k_0}) \text{ for some } 1 \le k_0 \le n\Big\}.$$

Since $U_i = T_i$ and the W_j's are all the distinct observations among $Z_1, \ldots, Z_n$, we know that

$$\Big[\# \text{ of } (U_i, W_j)\text{'}s \ \text{ satisfying } \ (U_i, W_j) = (T_{k_0}, Z_{k_0}) \ \text{ for some } \ 1 \le k_0 \le n\Big] = n,$$

which implies

$$\sum_{\substack{j=1 \\ \gamma_{ij}=1}}^{q} \sum_{i=1}^{n} I\{U_i \leq t, W_j \leq z\} \;=\; \sum_{i=1}^{n} I\{T_i \leq t, Z_i \leq z\}. \tag{2.3.49}$$

The proof follows from (2.3.48)-(2.3.49). $\square$

Proof of Lemma 2.1: Since $N_{ij} > 0$ for $(i,j) \in J_1$, where J_1 is given by (2.3.4), we know that any solution to (2.3.6) will satisfy $a_{ij} < 1$, which implies that optimization problem (2.3.6) is equivalent to

$$\begin{cases} \max \log G_1(\boldsymbol{a}_1) = \sum\sum_{(i,j)\in J_1} N_{ij} \log(1 - a_{ij}) \\ \text{subject to :} \quad 0 \leq a_{ij} < 1, \quad (i,j) \in J_1. \end{cases} \tag{2.3.50}$$

The proof follows from noting that $\log G_1(\boldsymbol{a}_1)$ is well-defined on set

$$A_1 \equiv \big\{\boldsymbol{a}_1 \,|\, 0 \leq a_{ij} < 1,\ (i,j) \in J_1\big\}, \tag{2.3.51}$$

and $\log G_1(\boldsymbol{a}_1)$ is a strictly decreasing function in each component of $\boldsymbol{a}_1 \in A_1$ because

$$\frac{\partial}{\partial a_{ij}}\Big(\log G_1(\boldsymbol{a}_1)\Big) = \frac{-N_{ij}}{1 - a_{ij}} < 0, \qquad (i,j) \in J_1. \qquad \square$$

Proof of Lemma 2.2: Since $\gamma_{ij} > 0$ for $(i,j) \in J_2$, where J_2 is given by (2.3.8), we know that any solution to (2.3.10) will satisfy $a_{ij} > 0$, which implies that optimization problem (2.3.10) is equivalent to

$$\begin{cases} \max \log G_2(\boldsymbol{a}_2) = \sum\sum_{(i,j)\in J_2} \gamma_{ij} \log a_{ij} \\ \text{subject to :} \quad 0 < a_{ij} \leq 1, \quad (i,j) \in J_2. \end{cases} \tag{2.3.52}$$

The proof follows from noting that $\log G_2(\boldsymbol{a}_2)$ is well-defined on set

$$A_2 \equiv \left\{\boldsymbol{a}_2 \,|\, 0 < a_{ij} \le 1,\ (i,j) \in J_2\right\}, \tag{2.3.53}$$

and $\log G_2(\boldsymbol{a}_2)$ is a strictly increasing function in each component of $\boldsymbol{a}_2 \in A_2$ because

$$\frac{\partial}{\partial a_{ij}}\Big(\log G_2(\boldsymbol{a}_2)\Big) = \frac{\gamma_{ij}}{a_{ij}} > 0, \qquad (i,j) \in J_2. \qquad \square$$

Proof of Lemma 2.3: Since $\gamma_{ij} > 0$ and $N_{ij} - \gamma_{ij} > 0$ for $(i,j) \in J_3$, where J_3 is given by (2.3.12), we know that any solution to (2.3.14) will satisfy $0 < a_{ij} < 1$, which implies that optimization problem (2.3.14) is equivalent to

$$\begin{cases} \max\ \log G_3(\boldsymbol{a}_3) = \displaystyle\sum\sum_{(i,j)\in J_3} \Big[\gamma_{ij}\log a_{ij} + (N_{ij} - \gamma_{ij})\log(1 - a_{ij})\Big] \\ \text{subject to}: \quad 0 < a_{ij} < 1, \quad (i,j) \in J_3. \end{cases} \tag{2.3.54}$$

Note that $\log G_3(\boldsymbol{a}_3)$ is well-defined on the convex set

$$A_3 \equiv \left\{\boldsymbol{a}_3 \,|\, 0 < a_{ij} < 1,\ (i,j) \in J_3\right\}. \tag{2.3.55}$$

Also, note that $\hat{a}_{ij} = \frac{\gamma_{ij}}{N_{ij}}$, $(i,j) \in J_3$, is a solution to

$$0 = \frac{\partial}{\partial a_{ij}}\Big(\log G_3(\boldsymbol{a}_3)\Big) = \frac{\gamma_{ij}}{a_{ij}} - \frac{N_{ij} - \gamma_{ij}}{1 - a_{ij}},$$

and $0 < \hat{a}_{ij} < 1$ since $\gamma_{ij} > 0$ and $N_{ij} - \gamma_{ij} > 0$ for $(i,j) \in J_3$. The proof follows from noting that for $(i,j), (k,\ell) \in J_3$, we have

$$\begin{cases} \dfrac{\partial^2}{\partial a_{ij}^2}\Big(\log G_3(\boldsymbol{a}_3)\Big) = \dfrac{-\gamma_{ij}}{(a_{ij})^2} - \dfrac{N_{ij} - \gamma_{ij}}{(1 - a_{ij})^2} \ < \ 0, \\ \dfrac{\partial^2}{\partial a_{ij}\partial a_{k\ell}}\Big(\log G_3(\boldsymbol{a}_3)\Big) = 0 \qquad\qquad \text{for } i \ne k \text{ or } j \ne \ell, \end{cases} \tag{2.3.56}$$

which implies that $\log G_3(\boldsymbol{a}_3)$ is strictly concave down on A_3. $\square$

Proof of Lemma 2.4: Note that from (2.2.19), J in (2.3.16) can be written as

$$J = \big\{(i,j) \,|\, N_{ij} > 0\big\} = J_1 \cup J_2 \cup J_3, \tag{2.3.57}$$

where $J_1 \cap J_2 \cap J_3 = \emptyset$ and J_1, J_2, and J_3 are given by (2.3.4), (2.3.8) and (2.3.12), respectively. Therefore, $G_4(\boldsymbol{a})$ in (2.3.17) can be written as

$$G_4(\boldsymbol{a}) = G_1(\boldsymbol{a}_1)\, G_2(\boldsymbol{a}_2)\, G_3(\boldsymbol{a}_3), \tag{2.3.58}$$

where $G_1(\boldsymbol{a}_1)$, $G_2(\boldsymbol{a}_2)$ and $G_3(\boldsymbol{a}_3)$ are given by (2.3.5), (2.3.9) and (2.3.13), respectively. Thus, from Lemmas 2.1-2.3 and (2.3.57), we know the solution to (2.3.18) is uniquely given by

$$\hat{\boldsymbol{a}} = \{\hat{a}_{ij} \,|\, (i,j) \in J\} = \hat{\boldsymbol{a}}_1 \cup \hat{\boldsymbol{a}}_2 \cup \hat{\boldsymbol{a}}_3,$$

where $\hat{\boldsymbol{a}}_1 = \mathbf{0}$ is the unique solution to (2.3.6), $\hat{\boldsymbol{a}}_2 = \mathbf{1}$ is the unique solution to (2.3.10) and $\hat{\boldsymbol{a}}_3$ is the unique solution to (2.3.14) and is given by (2.3.15). Therefore, we have

$$\begin{aligned}
\hat{a}_{ij} &= \begin{cases} 0, & \text{if } \ (i,j) \in J_1 \\ 1, & \text{if } \ (i,j) \in J_2 \\ \dfrac{\gamma_{ij}}{N_{ij}}, & \text{if } \ (i,j) \in J_3 \end{cases} \\
&= \begin{cases} 0, & \text{if } \ \gamma_{ij} = 0,\ N_{ij} - \gamma_{ij} > 0,\ 1 \le j \le q,\ 1 \le i \le m_j \\ 1, & \text{if } \ \gamma_{ij} > 0,\ N_{ij} - \gamma_{ij} = 0,\ 1 \le j \le q,\ 1 \le i \le m_j \\ \dfrac{\gamma_{ij}}{N_{ij}}, & \text{if } \ \gamma_{ij} > 0,\ N_{ij} - \gamma_{ij} > 0,\ 1 \le j \le q,\ 1 \le i \le m_j, \end{cases}
\end{aligned} \tag{2.3.59}$$

and the proof follows from (2.3.57) and (2.3.59). $\square$

Proof of Lemma 2.5: From $N_{1j} > 0$ in (2.2.19), we know that any solution to (2.3.21) will satisfy $b_{1j} > 0$, which implies, from $\sum_{j=1}^{q} b_{1j} = 1$, that optimization problem (2.3.21) is equivalent to

$$\begin{cases} \min\ -\log G_1(\boldsymbol{b}) = -\sum_{j=1}^{q} N_{1j} \log b_{1j} \\ \text{subject to :} \quad h(\boldsymbol{b}) = 0; \quad \boldsymbol{b} \in B, \end{cases} \tag{2.3.60}$$

where

$$h(\boldsymbol{b}) \equiv 1 - \sum_{j=1}^{q} b_{1j}, \qquad B \ \equiv \ \big\{\boldsymbol{b} \,|\, 0 < b_{1j} < 1,\ 1 \le j \le q\big\}. \tag{2.3.61}$$

As follows, we discuss the Karush-Kuhn-Tucker (KKT) sufficient conditions (Bazaraa *et al.*, 1993; page 164) for optimization problem (2.3.60).

Using Lagrange multipliers, let

$$\mathcal{G}(\boldsymbol{b}, \nu) \equiv -\log G_1(\boldsymbol{b}) + \nu h(\boldsymbol{b}) = -\sum_{j=1}^{q} N_{1j} \log b_{1j} + \nu\Big(1 - \sum_{j=1}^{q} b_{1j}\Big), \tag{2.3.62}$$

which has the following partial derivatives

$$\begin{cases} \dfrac{\partial}{\partial b_{1j}}\Big(\mathcal{G}(\boldsymbol{b}, \nu)\Big) = -\dfrac{N_{1j}}{b_{1j}} - \nu, & 1 \le j \le q \\ \dfrac{\partial}{\partial \nu}\Big(\mathcal{G}(\boldsymbol{b}, \nu)\Big) = 1 - \sum_{j=1}^{q} b_{1j}. \end{cases} \tag{2.3.63}$$

From $\partial G / \partial b_{1j} = 0$, we have

$$b_{1j} = \frac{-N_{1j}}{\nu}, \qquad 1 \le j \le q, \tag{2.3.64}$$

and from (2.2.6), (2.2.16) and $\partial G/\partial \nu = 0$, we have

$$1 = \sum_{j=1}^{q} b_{1j} = \frac{-1}{\nu} \sum_{j=1}^{q} N_{1j} = \frac{-1}{\nu} \sum_{j=1}^{q} \sum_{i=1}^{m} n_{ij} \ = \ \frac{-n}{\nu} \qquad \Longrightarrow \qquad \nu = -n. \tag{2.3.65}$$

Therefore, $\hat{b}_{1j} = \frac{N_{1j}}{n}$ is a solution to $\nabla \mathcal{G}(\boldsymbol{b}, \nu) = -\nabla \log G_1(\boldsymbol{b}) + \nu \nabla h(\boldsymbol{b}) = \mathbf{0}$.

Note that B in (2.3.61) is open and convex and that $\log G_1(\boldsymbol{b})$ and $h(\boldsymbol{b})$ are well-defined on B. We know $\hat{\boldsymbol{b}}$ is a feasible solution for (2.3.60) because from (2.3.64)-(2.3.65), we have $\sum_{j=1}^{q} \hat{b}_{1j} = 1$ and from (2.2.6) and (2.2.19), we have $0 < N_{1j} < n$, which implies $0 < \hat{b}_{1j} < 1$. Note that $-\log G_1(\boldsymbol{b})$ is strictly concave up on B because

$$\begin{cases} \dfrac{\partial^2}{\partial b_{1j}^2}\Big(-\log G_1(\boldsymbol{b})\Big) = \dfrac{N_{1j}}{(b_{1j})^2} \ > \ 0 \\ \dfrac{\partial^2}{\partial b_{1j} \partial b_{1k}}\Big(-\log G_1(\boldsymbol{b})\Big) = 0 \qquad \text{for} \ \ j \neq k. \end{cases} \tag{2.3.66}$$

Therefore, from Bazaraa *et al.* (1993; page 116), we know $-\log G_1(\boldsymbol{b})$ is pseudoconvex on B (see definition in Bazaraa *et al.*, 1993; page 113). Also, from Bazaraa *et al.* (1993; page 118, Problem 3.4, and page 116), we know $h(\boldsymbol{b})$ is both quasiconvex and quasiconcave (see definitions in Bazaraa *et al.*, 1993; page 108) since $h(\boldsymbol{b})$ is linear. The proof follows from the KKT sufficient conditions and Theorem 3.4.2 in Bazaraa *et al.* (1993; pages 164 and 101, respectively). $\square$

CHAPTER 3. EMPIRICAL LIKELIHOOD RATIO CONFIDENCE INTERVAL FOR CONDITIONAL SURVIVAL PROBABILITIES

In this chapter, we study the empirical likelihood ratio confidence interval for conditional survival probabilities with right censored bivariate data (1.2.23) or (2.1.2).

3.1 Introduction

In survival analysis, often interested is focused on the probability that a patient survives up to time t_0 given that the covariate Z is equal to a specified value z_0. For instance, one might be interested in the conditional survival probability given that the patient is a male, or the conditional survival probability given that the patient received a particular treatment. In this chapter, we consider right censored data (1.2.23) or (2.1.2), where Z is a discrete variable and z_0 is one of the possible values of Z, and we construct the empirical likelihood-based confidence interval for the following conditional probability:

$$\theta_0 = P\{T \leq t_0 \,|\, Z = z_0\}. \tag{3.1.1}$$

Note that the confidence interval for conditional survival probability

$$(1 - \theta_0) = P\{T > t_0 \,|\, Z = z_0\} \tag{3.1.2}$$

is equivalent to that for θ_0.

In Section 3.2, we show that the empirical likelihood ratio confidence set for θ_0 is in fact an interval. To study the asymptotic behavior of the empirical likelihood ratio, the expression of such ratio is needed and is a rather complex issue. We provide an analytic solution for such likelihood ratio in Section 3.2. Section 3.3 discusses the computation of the empirical likelihood ratio confidence interval (ELRCI) for θ_0. All proofs are deferred to Sections 3.4-3.5.

3.2 Confidence Interval for Conditional Survival Probabilities

Note that from (2.2.1) we have that as $n \to \infty$

$$U_m > t_0 \qquad \text{in probability,} \tag{3.2.1}$$

because as $n \to \infty$, we have that for d.f. F_V of V with $0 < F_V(t_0) < 1$,

$$P\big\{U_m \le t_0\big\} = P\big\{V_1 \le t_0, \cdots, V_n \le t_0\big\} = \big(F_V(t_0)\big)^n \to 0.$$

Also, note that $0 < P\{Z = z_0\} < 1$ since Z is discrete and z_0 is one the possible values of Z, which implies that as $n \to \infty$

$$\exists\ 1 \le \zeta \le q \ \text{ such that } \ W_\zeta = z_0 \qquad \text{in probability,} \tag{3.2.2}$$

where the proof of (3.2.2) is similar to that of (3.2.1) above. Let $F(t, z)$, $\boldsymbol{p} = (p_{ij})$, and $\boldsymbol{q} = (q_{ij})$ be defined as in (2.2.12)-(2.2.14), where for p_{ij}'s and q_{ij}'s we use the same treatment as in (2.2.18)-(2.2.19) for $\hat{p}_{ij}$'s and $\hat{q}_{ij}$'s. Then, we have

$$\begin{cases} F(t,z) = \sum_{j=1}^{q}\sum_{i=1}^{m} q_{ij} I\{U_i \le t, W_j \le z\}, \quad \text{for } \ t \le U_m, \quad z \in \mathbb{R} \\ p_{ij} = q_{ij}, \qquad \text{for } \ 1 \le j \le q, \ 1 \le i \le m_j \\ q_{ij} = 0, \qquad \text{for } \ 1 \le j \le q, \ m_j < i \le m \\ q_{m+1,j} = P_F\{T > U_m, \ Z = W_j\}, \quad \text{for } \ 1 \le j \le q \\ p_{m_j+1,j} = P_F\{T > U_{m_j}, \ Z = W_j\} = q_{m+1,j}, \quad \text{for } \ 1 \le j \le q \\ \sum_{j=1}^{q}\sum_{i=1}^{m_j+1} p_{ij} = \sum_{j=1}^{q}\sum_{i=1}^{m+1} q_{ij} = 1. \end{cases} \tag{3.2.3}$$

If we let $T(F)$ be the following statistical functional

$$T(F) = P_F\{T \le t_0 \,|\, Z = z_0\},$$

then from (3.2.1)-(3.2.3) we have that in probability as $n \to \infty$,

$$\begin{aligned} T(F) &= \frac{P_F\{T \le t_0, \ Z = z_0\}}{P_F\{Z = z_0\}} \\ &= \frac{P_F\{T \le t_0, \ Z = W_\zeta\}}{P_F\{T \le U_m, \ Z = W_\zeta\} + P_F\{T > U_m, \ Z = W_\zeta\}} \\ &= \frac{\sum_{j=1}^{q}\sum_{i=1}^{m} q_{ij} I\{U_i \le t_0, \ W_j = W_\zeta\}}{\sum_{j=1}^{q}\sum_{i=1}^{m} q_{ij} I\{U_i \le U_{m_j}, \ W_j = W_\zeta\} + q_{m+1,\zeta}} \\ &= \frac{\sum_{j=1}^{q}\sum_{i=1}^{m_j} p_{ij} I\{U_i \le t_0, \ W_j = W_\zeta\}}{\sum_{j=1}^{q}\sum_{i=1}^{m_j} p_{ij} I\{U_i \le U_{m_j}, \ W_j = W_\zeta\} + p_{m_\zeta+1,\zeta}} \\ &= \frac{\sum_{i=1}^{m_\zeta} p_{i\zeta} I\{U_i \le t_0\}}{\sum_{i=1}^{m_\zeta} p_{i\zeta} + p_{m_\zeta+1,\zeta}} = \frac{\sum_{i=1}^{m_\zeta} p_{i\zeta} I\{U_i \le t_0\}}{\sum_{i=1}^{m_\zeta+1} p_{i\zeta}} \equiv T(\boldsymbol{p}). \end{aligned} \tag{3.2.4}$$

For $L(F)$ given by (2.2.11) and $\hat{F}_n$ given by (2.2.18), we let

$$R(F) \equiv \frac{L(F)}{L(\hat{F}_n)} = \frac{L(\boldsymbol{p})}{L(\hat{F}_n)}. \tag{3.2.5}$$

Thus, from (3.2.4)-(3.2.5), the *empirical likelihood ratio* is denoted by

$$r(\theta_0) \equiv \sup_F \Big\{ R(F) \,\Big|\, T(F) = \theta_0 \Big\} = \frac{1}{L(\hat{F}_n)} \sup_{\boldsymbol{p}} \Big\{ L(\boldsymbol{p}) \,\Big|\, T(\boldsymbol{p}) = \theta_0,\ \boldsymbol{p} \in \mathcal{F}_n \Big\}, \tag{3.2.6}$$

where

$$\mathcal{F}_n = \left\{ \boldsymbol{p} \,\middle|\, 0 \le p_{ij} \le 1,\ 1 \le j \le q,\ 1 \le i \le m_j;\ \sum_{j=1}^{q} \sum_{i=1}^{m_j+1} p_{ij} = 1 \right\}. \tag{3.2.7}$$

From (1.4.22)-(1.4.22a), we see that for $0 < c < 1$, the confidence set S_n for the conditional probability θ_0 based on right censored data (1.2.23) or (2.1.2) is given by

$$\begin{aligned} S_n &= \big\{ \theta \,|\, r(\theta) \ge c \big\} = \big\{ T(F) \,\big|\, R(F) \ge c \big\} \\ &= \big\{ T(\boldsymbol{p}) \,|\, L(\boldsymbol{p}) \ge cL(\hat{F}_n),\ \boldsymbol{p} \in \mathcal{F}_n \big\} = \big\{ T(\boldsymbol{p}) \,|\, \boldsymbol{p} \in \mathcal{E}_n \big\}, \end{aligned} \tag{3.2.8}$$

where

$$\mathcal{E}_n = \left\{ \boldsymbol{p} \,\middle|\, L(\boldsymbol{p}) \ge cL(\hat{F}_n),\ \boldsymbol{p} \in \mathcal{F}_n \right\}. \tag{3.2.9}$$

In Section 3.4, we prove the following theorems on the confidence set S_n for θ_0.

Theorem 3.1. *S_n given by (3.2.8) is an interval satisfying $S_n = [T_L, T_U]$, where*

$$T_L = \inf_{\boldsymbol{p} \in \mathcal{E}_n} T(\boldsymbol{p}) \qquad \text{and} \qquad T_U = \sup_{\boldsymbol{p} \in \mathcal{E}_n} T(\boldsymbol{p}). \tag{3.2.10}$$

Theorem 3.2. *For S_n and $r(\theta_0)$ given by (3.2.8) and (3.2.6), respectively, we have that in probability as $n \to \infty$,*

$$\theta_0 \in S_n \quad \text{if and only if} \quad r(\theta_0) \ge c. \tag{3.2.11}$$

Note that (3.2.4) and Theorems 3.1-3.2 imply

$$P\{T_L \le \theta_0 \le T_U\} = P\{-2\log r(\theta_0) \le -2\log c\} + o_p(1). \tag{3.2.12}$$

Thus, the asymptotic behavior of empirical likelihood ratio confidence interval (ELRCI) $[T_L, T_U]$ can be studied via the empirical log likelihood ratio $\log r(\theta_0)$. An analytic solution for $\log r(\theta_0)$ based on data (1.2.23) or (2.1.1) is given in the following theorem with the proof deferred to Section 3.4, while the computation of T_L and T_U is discussed in Section 3.3.

Theorem 3.3. *Under the following conditions:*

$$\Lambda_n \equiv \min\{N_{i\zeta} - \gamma_{i\zeta} \,|\, 1 \le i \le m_\zeta,\, U_i \le t_0\} > 0, \quad \text{in probability} \tag{AS3.1}$$

$$\prod_{\substack{i=1\\U_i\le t_0}}^{m_\zeta}\Big(1 - \frac{\gamma_{i\zeta}}{N_{i\zeta} - \Lambda_n}\Big) < 1 - \theta_0 < \prod_{\substack{i=1\\U_i\le t_0}}^{m_\zeta}\Big(1 - \frac{\gamma_{i\zeta}}{N_{i\zeta}}\Big), \quad \text{in probability} \tag{AS3.2}$$

an expression of (3.2.6) is given by

$$\log r(\theta_0) = \sum_{\substack{i=1\\U_i\le t_0}}^{m_\zeta}\left[N_{i\zeta}\log\frac{N_{i\zeta}}{N_{i\zeta} - \tilde\lambda} + (N_{i\zeta} - \gamma_{i\zeta})\log\frac{N_{i\zeta} - \tilde\lambda - \gamma_{i\zeta}}{N_{i\zeta} - \gamma_{i\zeta}}\right], \tag{3.2.13}$$

in probability, where in probability $\tilde\lambda \in (0, \Lambda_n)$ is the unique root of

$$g(\lambda) \equiv \log(1-\theta_0) - \sum_{\substack{i=1\\U_i\le t_0}}^{m_\zeta}\log\Big(1 - \frac{\gamma_{i\zeta}}{N_{i\zeta} - \lambda}\Big). \tag{3.2.14}$$

Remark 3.1: It is expected that $-2\log r(\theta_0)$ converges in distribution to a Chi-squared distribution with 1 degree of freedom as $n \to \infty$. This will be further studied, and our result (3.2.13) here will facilitate this future work. For simulation studies on $-2\log r(\theta_0)$, above equation (3.2.14) may be solved using, say, Newton-Raphson method.

Remark 3.2: The meaning of Assumption (AS3.1) in Theorem 3.3 may be understood as follows. From (2.2.16) and (3.2.1), we know that in probability, for any $1 \leq i \leq m_\zeta$ satisfying $U_i \leq t_0$,

$$\frac{N_{i\zeta}}{n} \geq P\{V \geq t_0,\, Z = z_0\} + o_p(1) \tag{3.2.15}$$

$$\frac{\gamma_{i\zeta}}{n} \leq P\{V \geq t_0,\, \delta = 1,\, Z = z_0\} + o_p(1). \tag{3.2.16}$$

Thus, for any $1 \leq i \leq m_\zeta$ satisfying $U_i \leq t_0$ we have in probability

$$\frac{N_{i\zeta} - \gamma_{i\zeta}}{n} \geq P\{V \geq t_0,\, \delta = 0,\, Z = z_0\} + o_p(1), \tag{3.2.17}$$

which implies in (AS3.1) we have $\Lambda_n/n > 0$ in probability, provided $P\{V \geq t_0,\, \delta = 0,\, Z = z_0\} > 0$. As for Assumption (AS3.2), one should notice that the last term of inequality is the conditional Kaplan-Meier estimation (Kaplan and Meier, 1958) for $(1 - \theta_0)$. Thus, with (AS3.1) we have non-strict inequalities hold in probability in (AS3.2). The strict inequalities are required only for the unique existence of the solution for $g(\lambda) = 0$ in (3.2.14), while it is shown in Section 3.5 that $g(\lambda)$ is strictly increasing.

3.3 Computation of Confidence Interval

In this section, we discuss the computation of T_L and T_U in (3.2.10) for the ELRCI in Theorem 3.1. In particular, we outline the details for finding the lower bound T_L for the ELRCI, while the upper bound T_U for the ELRCI will be studied further in the future.

To find an expression for T_L in (3.2.10), we solve the following optimization problem

$$T_L = \begin{cases} \min\ T(\boldsymbol{p}) = \dfrac{\sum_{i=1}^{m_\zeta} p_{i\zeta} I\{U_i \leq t_0\}}{\sum_{i=1}^{m_\zeta+1} p_{i\zeta}} \\ \text{subject to}: \quad 0 \leq p_{ij} \leq 1, \quad \text{for} \ \ 1 \leq j \leq q,\ 1 \leq i \leq m_j; \\ \qquad\qquad\qquad L(\boldsymbol{p}) \geq cL(\hat{F}_n); \quad \sum_{j=1}^{q} \sum_{i=1}^{m_j+1} p_{ij} = 1. \end{cases} \tag{3.3.1}$$

Consider the transformation $\tau(\boldsymbol{p}) = (\boldsymbol{a}, \boldsymbol{b})$, where $\boldsymbol{a} = \{a_{ij} \,|\, 1 \le j \le q,\ 1 \le i \le m_j\}$ and $\boldsymbol{b} = (b_{11}, \cdots, b_{1q})$ are given by (2.3.23). Under this transformation, we have from (3.2.4) and (2.3.32)

$$\begin{aligned}
T(\boldsymbol{p}) &= \frac{\sum_{i=1}^{m_\zeta} a_{i\zeta} b_{1\zeta} \prod_{k=1}^{i-1}(1 - a_{k\zeta}) I\{U_i \le t_0\}}{b_{1\zeta}} \\
&= \sum_{i=1}^{m_\zeta} a_{i\zeta} \prod_{k=1}^{i-1}(1 - a_{k\zeta}) I\{U_i \le t_0\} \\
&= \sum_{i=1}^{m_\zeta} \Big(1 - (1 - a_{i\zeta})\Big) I\{U_i \le t_0\} \prod_{k=1}^{i-1}(1 - a_{k\zeta}) \\
&= \sum_{i=1}^{m_\zeta} \left(\prod_{k=1}^{i-1}(1 - a_{k\zeta}) - \prod_{k=1}^{i}(1 - a_{k\zeta}) \right) I\{U_i \le t_0\} \\
&= 1 - \prod_{i=1}^{m_\zeta}(1 - a_{i\zeta}) I\{U_i \le t_0\} \ = \ 1 - \prod_{\substack{i=1 \\ U_i \le t_0}}^{m_\zeta} (1 - a_{i\zeta}) = 1 - T_1(\boldsymbol{a}),
\end{aligned} \tag{3.3.2}$$

where

$$T_1(\boldsymbol{a}) \equiv \prod_{\substack{i=1 \\ U_i \le t_0}}^{m_\zeta} (1 - a_{i\zeta}). \tag{3.3.3}$$

From (2.3.28) and (3.3.3), we see that optimization problem (3.3.1) is equivalent to

$$T_L = \begin{cases} \min\ 1 - T_1(\boldsymbol{a}) \\ \text{subject to :} \quad 0 \le a_{ij} \le 1,\ 0 \le b_{1j} \le 1,\ (i,j) \in J; \\ \qquad\qquad\qquad G(\boldsymbol{a}, \boldsymbol{b}) \ge cL(\hat{F}_n); \quad \sum_{j=1}^{q} b_{1j} = 1, \end{cases}$$

where J is given by (2.3.16), which implies that

$$T_L = 1 - T_1(\boldsymbol{a}^L), \tag{3.3.4}$$

where

$$
-T_1(\boldsymbol{a}^L) = \begin{cases} \min \ -T_1(\boldsymbol{a}) \\ \text{subject to}: \quad 0 \le a_{ij} \le 1,\ 0 \le b_{1j} \le 1,\ (i,j) \in J; \\ \qquad\qquad\qquad G(\boldsymbol{a}, \boldsymbol{b}) \ge cL(\hat{F}_n); \quad \sum_{j=1}^{q} b_{1j} = 1. \end{cases} \tag{3.3.5}
$$

As follows, we discuss the solution to (3.3.5).

If we define

$$
\begin{cases} g_1(\boldsymbol{a}, \boldsymbol{b}) \equiv \log\big(cL(\hat{F}_n)\big) - \log G(\boldsymbol{a}, \boldsymbol{b}) \\ h_1(\boldsymbol{b}) \equiv 1 - \sum_{j=1}^{q} b_{1j} \\ X_1 \equiv \big\{\boldsymbol{a} \,\big|\, 0 < a_{ij} < 1,\ (i,j) \in J\big\} \\ X_2 \equiv \big\{\boldsymbol{b} \,\big|\, 0 < b_{1j} < 1,\ 1 \le j \le q\big\}, \end{cases} \tag{3.3.6}
$$

then to find the solution to optimization problem (3.3.5), we solve

$$
\begin{cases} \min \ -\log T_1(\boldsymbol{a}) \\ \text{subject to}: \quad (\boldsymbol{a}, \boldsymbol{b}) \in X_1 \times X_2; \quad g_1(\boldsymbol{a}, \boldsymbol{b}) \le 0; \quad h_1(\boldsymbol{b}) = 0. \end{cases} \tag{3.3.7}
$$

Using Lagrange multipliers with $\lambda_1^L \ge 0$, let

$$
\begin{aligned} H_L(\boldsymbol{a}, \boldsymbol{b}, \lambda_1, \lambda_2) &\equiv -\log T_1(\boldsymbol{a}) + \lambda_1 g_1(\boldsymbol{a}, \boldsymbol{b}) + \lambda_2 h_1(\boldsymbol{b}) \\ &= -\sum_{(i\zeta) \in \mathcal{J}_1} \log(1 - a_{k\zeta}) + \lambda_2 \Big(1 - \sum_{j=1}^{q} b_{1j}\Big) \\ &\quad + \lambda_1 \Big\{ \log\big(cL(\hat{F}_n)\big) - \sum_{j=1}^{q} N_{1j} \log b_{1j} \\ &\qquad - \sum\sum_{(i,j) \in J} \Big[\gamma_{ij} \log a_{ij} + (N_{ij} - \gamma_{ij}) \log(1 - a_{ij})\Big] \Big\}, \end{aligned} \tag{3.3.8}
$$

where $\mathcal{J}_1$ is given in (3.5.1), and note that H_L has the following partial derivatives

$$\begin{cases} \dfrac{\partial H_L}{\partial a_{ij}} = \dfrac{1}{1-a_{ij}} I\big\{(i,\zeta) \in \mathcal{J}_1\big\} - \lambda_1\Big(\dfrac{\gamma_{ij}}{a_{ij}} - \dfrac{N_{ij}-\gamma_{ij}}{1-a_{ij}}\Big), \qquad (i,j) \in J \\ \dfrac{\partial H_L}{\partial b_{1j}} = -\lambda_2 - \lambda_1\Big(\dfrac{N_{1j}}{b_{1j}}\Big), \qquad 1 \le j \le q \\ \dfrac{\partial H_L}{\partial \lambda_1} = g_1(\boldsymbol{a}, \boldsymbol{b}) \\ \dfrac{\partial H_L}{\partial \lambda_2} = h_1(\boldsymbol{b}). \end{cases} \tag{3.3.9}$$

From $\partial H_L/\partial a_{ij} = 0$ in (3.3.9), we have

$$a_{ij} = \begin{cases} \dfrac{\lambda_1 \gamma_{ij}}{\lambda_1 N_{ij} + 1} & \text{for} \quad (i,j) \in \mathcal{J}_1 \\ \dfrac{\gamma_{ij}}{N_{ij}} & \text{for} \quad (i,j) \in \mathcal{J}_2. \end{cases} \tag{3.3.10}$$

where $\mathcal{J}_2$ is given by (3.5.5) and we note that $\mathcal{J}_1 \cup \mathcal{J}_2 = J$. From $\partial H_L/\partial b_{1j} = 0$ in (3.3.9), we have

$$b_{1j} = \frac{-\lambda_1}{\lambda_2} N_{1j}, \quad 1 \le j \le q, \tag{3.3.11}$$

where from $\partial H_L/\partial \lambda_2 = h_1(\boldsymbol{b}) = 0$ in (3.3.9), we have from (3.3.6), (3.3.11), (2.2.6) and (2.2.16)

$$1 = \sum_{j=1}^{q} b_{1j} = \frac{-\lambda_1}{\lambda_2} \sum_{j=1}^{q} N_{1j} = \frac{-\lambda_1}{\lambda_2} \sum_{j=1}^{q} \sum_{i=1}^{m} n_{ij} = \frac{-\lambda_1 n}{\lambda_2} \qquad \Longrightarrow \qquad \lambda_2 = -\lambda_1 n \tag{3.3.12}$$

From $\partial H_L/\partial \lambda_1 = 0$ in (3.3.9), we have from (3.3.6)

$$0 = g_1(\boldsymbol{a}, \boldsymbol{b}) = \log\big(cL(\hat{F}_n)\big) - \log G(\boldsymbol{a}, \boldsymbol{b}) \; = \; \log c + \log \frac{L(\hat{F}_n)}{G(\boldsymbol{a}, \boldsymbol{b})}. \tag{3.3.13}$$

Let $(\boldsymbol{a}^L, \boldsymbol{b}^L)$ be a solution to equations (3.3.9)-(3.3.13). Thus, we know

$$g_1(\boldsymbol{a}^L, \boldsymbol{b}^L) = 0; \qquad h_1(\boldsymbol{b}^L) = 0; \tag{3.3.14}$$

$$\nabla - \log T_1(\boldsymbol{a}^L) + \lambda_1 \nabla g_1(\boldsymbol{a}^L, \boldsymbol{b}^L) + \lambda_2 \nabla h_1(\boldsymbol{b}^L) = \nabla H_L(\boldsymbol{a}^L, \boldsymbol{b}^L, \lambda_1^L, \lambda_2^L) = \mathbf{0}, \tag{3.3.15}$$

with $\boldsymbol{a}^L = \{a_{ij}^L \,|\, (i,j) \in J\}$ and $\boldsymbol{b}^L = (b_{11}^L, \cdots, b_{1q}^L)$, where

$$a_{ij}^L = \begin{cases} \dfrac{\lambda_1^L \gamma_{ij}}{\lambda_1^L N_{ij} + 1} & \text{for } \ (i,j) \in \mathcal{J}_1 \\ \dfrac{\gamma_{ij}}{N_{ij}} & \text{for } \ (i,j) \in \mathcal{J}_2 \end{cases} \tag{3.3.16}$$

$$b_{1j}^L = \frac{N_{1j}}{n}, \quad 1 \le j \le q \tag{3.3.17}$$

and λ_1^L is a solution to $g_2(\lambda) = 0$, where

$$g_2(\lambda) \equiv \log c + \frac{L(\hat{F}_n)}{G(\boldsymbol{a}^L, \boldsymbol{b}^L)}. \tag{3.3.18}$$

Note that from (2.3.28) and (2.3.30), we know $L(\hat{F}_n) = G(\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}})$ with $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$ given by (2.3.19) and (2.3.22), respectively. Noting that $a_{ij}^L = \hat{a}_{ij}$ for $(i,j) \in \mathcal{J}_2$ and $b_{1j}^L = \hat{b}_{1j}$ for $1 \le j \le q$, we have from (3.3.16), (3.3.18), (2.3.19) and (2.3.28)

$$\begin{aligned} g_2(\lambda) &= \log c + \log \frac{\left(\prod_{j=1}^{q} (\hat{b}_{1j})^{N_{1j}}\right) \prod\prod_{(i,j)\in J} (\hat{a}_{ij})^{\gamma_{ij}} \left(1 - \hat{a}_{ij}\right)^{N_{ij} - \gamma_{ij}}}{\left(\prod_{j=1}^{q} (b_{1j}^L)^{N_{1j}}\right) \prod\prod_{(i,j)\in J} (a_{ij}^L)^{\gamma_{ij}} \left(1 - a_{ij}^L\right)^{N_{ij} - \gamma_{ij}}} \\ &= \log c + \log \prod\prod_{(i,j)\in\mathcal{J}_1} \left(\frac{\hat{a}_{ij}}{a_{ij}^L}\right)^{\gamma_{ij}} \left(\frac{1 - \hat{a}_{ij}}{1 - a_{ij}^L}\right)^{N_{ij} - \gamma_{ij}} \\ &= \log c + \sum\sum_{(i,j)\in\mathcal{J}_1} \left[\gamma_{ij} \log \frac{\hat{a}_{ij}}{a_{ij}^L} + (N_{ij} - \gamma_{ij}) \log \frac{1 - \hat{a}_{ij}}{1 - a_{ij}^L}\right] \\ &= \log c + \sum_{\substack{i=1 \\ U_i \le t_0}}^{m_\zeta} \left[\gamma_{i\zeta} \log \frac{\lambda N_{i\zeta} + 1}{\lambda N_{i\zeta}} + (N_{i\zeta} - \gamma_{i\zeta}) \log \frac{(N_{i\zeta} - \gamma_{i\zeta})(\lambda N_{i\zeta} + 1)}{N_{i\zeta}(\lambda N_{i\zeta} - \lambda\gamma_{i\zeta} + 1)}\right]. \end{aligned} \tag{3.3.19}$$

As follows, we discuss the Karush-Kuhn-Tucker sufficient conditions (Bazaraa *et al.*, 1993; page 164) for optimization problem (3.3.7).

Note that $-\log T_1$, g_1 and h_1 in (3.3.6) are well-defined on the open, convex set $X_1 \times X_2$. Also, note that $-\log T_1(\boldsymbol{a})$ is strictly is concave up on X_1 because for all $\boldsymbol{a} \in X_1$, we have

$$\frac{\partial^2}{\partial a_{i\zeta}^2}\Big(-\log T_1(\boldsymbol{a})\Big) = \frac{1}{(1-a_{i\zeta})^2} > 0$$
$$\frac{\partial^2}{\partial a_{i\zeta} a_{k\zeta}}\Big(-\log T_1(\boldsymbol{a})\Big) = 0 \qquad \text{for} \quad i \neq k.$$

Therefore, from Bazaraa *et al.* (1993; page 116), we know $-\log T_1(\boldsymbol{a})$ is pseudoconvex on X_1 (see definition in Bazaraa *et al.*, 1993; page 113). For X given by (3.4.8), we know $(\boldsymbol{a}^L, \boldsymbol{b}^L) \in X$, which implies from (3.4.13) that $-\log G(\boldsymbol{a}, \boldsymbol{b})$ is quasiconvex at $(\boldsymbol{a}^L, \boldsymbol{b}^L)$, in turn, $g_1(\boldsymbol{a}, \boldsymbol{b})$ in (3.3.6) is quasiconvex at $(\boldsymbol{a}^L, \boldsymbol{b}^L)$ (see definition in Bazaraa *et al.*, 1993; page 108). Also, from Bazaraa *et al.* (1993; page 118, Problem 3.4, and page 116), we know $h_1(\boldsymbol{b})$ is both quasiconvex and quasiconcave on X_2 (see definitions in Bazaraa *et al.*, 1993; page 108) since $h_1(\boldsymbol{b})$ is linear.

Note that we can relax the restriction on $\boldsymbol{a}$ in $X_1 \times X_2$ given by (3.3.6) to

$$\begin{cases} 0 \leq a_{ij} < 1, & \text{if} \quad \gamma_{ij} = 0, \quad (i,j) \in \mathcal{J}_1 \\ 0 < a_{ij} < 1, & \text{otherwise} \end{cases} \tag{3.3.20}$$

because if $\gamma_{ij} = 0$, $(i,j) \in \mathcal{J}_1$, then $a_{ij}^L = 0$ in (3.3.16) and we have in the sum for $G(\boldsymbol{a}, \boldsymbol{b})$ in (2.3.28):

$$\gamma_{ij} \log a_{ij} = 0 \log 0 = 0,$$

thus $g_1(\boldsymbol{a}, \boldsymbol{b})$ in (3.3.6) is well-defined. Noting that $\gamma_{ij} \geq 0$ for all $(i,j) \in J$, we see that $\boldsymbol{a}^L$ in (3.3.16) satisfies (3.3.20) for all $(i,j) \in J$ and $\lambda_1^L > 0$, which implies $(\boldsymbol{a}^L, \boldsymbol{b}^L)$ is a feasible solution for the minimization problem in (3.3.7) if $\lambda_1^L > 0$ is as solution to $g_2(\lambda) = 0$.

Therefore, from the KKT conditions and Theorem 3.4.2 in Bazaraa *et al.* (1993; pages 164 and 101), the solution to minimization problem (3.3.7) is uniquely given by $(\boldsymbol{a}^L, \boldsymbol{b}^L)$ in (3.3.16)-(3.3.17) if $\lambda_1^L > 0$ is the unique solution to $g_2(\lambda) = 0$ in (3.3.19). As follows, we discuss the unique existence of the solution to $g_2(\lambda) = 0$ on $(0, \infty)$.

Note that for $\lambda > 0$, we have from (3.3.19) and Lemma 3.1 (ii) that as $n \to \infty$,

$$\begin{aligned} g_2'(\lambda) &= \sum_{\substack{i=1 \\ U_i \le t_0}}^{m_\zeta} \left[\frac{-\gamma_{i\zeta}}{\lambda(\lambda N_{i\zeta} + 1)} + \frac{\gamma_{i\zeta}(N_{i\zeta} - \gamma_{i\zeta})}{(\lambda N_{i\zeta} + 1)(\lambda N_{i\zeta} - \lambda\gamma_{i\zeta} + 1)} \right] \\ &= \sum_{\substack{i=1 \\ U_i \le t_0}}^{m_\zeta} \frac{-\gamma_{i\zeta}}{\lambda(\lambda N_{i\zeta} + 1)(\lambda N_{i\zeta} - \lambda\gamma_{i\zeta}) + 1)} < 0, \quad \text{in probability}, \end{aligned} \tag{3.3.21}$$

which implies that $g_2(\lambda)$ is a strictly decreasing function on $(0, \infty)$. Also, note that if we let

$$k_1 \equiv \log c + \sum_{\substack{i=1 \\ U_i \le t_0}}^{m_\zeta} (N_{i\zeta} - \gamma_{i\zeta}) \log \frac{N_{i\zeta} - \gamma_{i\zeta}}{N_{i\zeta}}, \tag{3.3.22}$$

then we have from (3.3.19),

$$\begin{aligned} \lim_{\lambda \to 0+} g_2(\lambda) &= \log c + \sum_{\substack{i=1 \\ U_i \le t_0}}^{m_\zeta} \left[-\gamma_{i\zeta} \lim_{\lambda \to 0+} \log(\lambda N_{i\zeta}) + (N_{i\zeta} - \gamma_{i\zeta}) \log \frac{N_{i\zeta} - \gamma_{i\zeta}}{N_{i\zeta}} \right] \\ &= k_1 - \sum_{\substack{i=1 \\ U_i \le t_0}}^{m_\zeta} \gamma_{i\zeta} \lim_{\lambda \to 0+} \log(\lambda N_{i\zeta}) = \infty \end{aligned} \tag{3.3.23}$$

and since $0 < c < 1$, we have

$$\lim_{\lambda \to \infty} g_2(\lambda) = \log c < 0. \tag{3.3.24}$$

Since $g_2(\lambda)$ is well-defined on $(0, \infty)$, we see that $g_2(\lambda)$ is continuous on $(0, \infty)$. Therefore, from the Intermediate Value Theorem, we see that there exists a unique solution to $g_2(\lambda) = 0$ on $(0, \infty)$.

Therefore, the lower bound of the ELRCI is given by (3.3.4) with $\boldsymbol{a}^L$ given by (3.3.16), where λ_1^L is the solution to $g_2(\lambda) = 0$ on $(0, \infty)$. Note that we may use the Newton-Raphson method to find this solution.

3.4 Proofs of Theorems 3.1-3.2

Proof of Theorem 3.1: Note that set $\mathcal{E}_n$ given by (3.2.9) is compact because set $\mathcal{F}_n$ in (3.2.7) is compact and $L(\boldsymbol{p})$ in (2.2.11) is a polynomial, thus is continuous in $\boldsymbol{p}$. Also, note that $T(\boldsymbol{p})$ in (3.2.4) is well-defined on $\mathcal{E}_n$ in (3.2.9) because

$$\boldsymbol{p} \in \mathcal{E}_n \quad \Longrightarrow \quad \sum_{k=i}^{m_j+1} p_{kj} > 0, \qquad \text{for all } 1 \le j \le q,\ 1 \le i \le m_j, \tag{3.4.1}$$

which implies that the denominator of $T(\boldsymbol{p})$ in (3.2.4) is positive for any $\boldsymbol{p} \in \mathcal{E}_n$; i.e,

$$\boldsymbol{p} \in \mathcal{E}_n \quad \Longrightarrow \quad \sum_{k=1}^{m_j+1} p_{kj} > 0. \tag{3.4.2}$$

To see (3.4.1), it suffices to notice that for any $\boldsymbol{p} \in \mathcal{E}_n$, we have

$$L(\boldsymbol{p}) \ge cL(\hat{F}_n) > 0, \tag{3.4.3}$$

which implies the following in the product of $L(\boldsymbol{p})$ in (2.2.11):

$$\begin{aligned} 0 &< (p_{m_j,j})^{\gamma_{m_j,j}} (p_{m_j+1,j})^{n_{m_j,j} - \gamma_{m_j,j}} \\ &= \begin{cases} (p_{m_j+1,j})^{n_{m_j,j}} & \text{if} \quad \gamma_{m_j,j} = 0 \\ (p_{m_j,j})^{\gamma_{m_j,j}} (p_{m_j+1,j})^{n_{m_j,j} - \gamma_{m_j,j}} & \text{if} \quad \gamma_{m_j,j} > 0 \end{cases} \end{aligned}$$

in turn, from $n_{m_j,j} > 0$ in (2.2.19), we have $p_{m_j+1,j} > 0$ when $\gamma_{m_j,j} = 0$; $p_{m_j,j} > 0$ when $\gamma_{m_j,j} > 0$; which give $p_{m_j,j} + p_{m_j+1,j} > 0$.

Since the numerator of $T(\boldsymbol{p})$ in (3.2.4) is a linear function in $\boldsymbol{p}$, from (3.4.2) we know that $T(\boldsymbol{p})$ is continuous on $\mathcal{E}_n$. Thus, from Royden (1988; page 191), we know that S_n given by (3.2.8) is a compact set in $\mathbb{R}$. Note that if $\mathcal{E}_n$ is connected, then S_n is connected (Royden, 1988; page 182), which implies that S_n is either an interval or a single point (Royden, 1988; page 183). Since S_n is compact, we know that S_n is a closed interval $[T_L, T_U]$ with T_L and T_U given by (3.2.10). Thus, if we define

$$\tau(\boldsymbol{p}) = (\boldsymbol{a}, \boldsymbol{b}) \quad \Longleftrightarrow \quad \begin{cases} a_{ij} = \dfrac{p_{ij}}{b_{ij}} \\ b_{ij} = \sum_{k=i}^{m_j+1} p_{kj} \end{cases} \qquad 1 \le j \le q, \quad 1 \le i \le m_j, \tag{3.4.4}$$

where $\boldsymbol{a} = (a_{ij})$ and $\boldsymbol{b} = (b_{1j})$, the proof follows from showing: (I) $\tau(\mathcal{E}_n)$ is convex; (II) $\mathcal{E}_n$ is connected; which are proved as follows.

(I) "$\boldsymbol{\tau(\mathcal{E}_n)}$ is convex": Note that from (3.4.1), τ is well-defined on $\mathcal{E}_n$. Also note that under transformation (3.4.4), we have from (2.3.28)

$$L(\boldsymbol{p}) = L\Big(\tau^{-1}(\boldsymbol{a}, \boldsymbol{b})\Big) = \left(\prod_{j=1}^{q} (b_{1j})^{N_{1j}}\right) \prod_{j=1}^{q} \prod_{i=1}^{m_j} (a_{ij})^{\gamma_{ij}} (1 - a_{ij})^{N_{ij} - \gamma_{ij}} = G(\boldsymbol{a}, \boldsymbol{b}), \tag{3.4.5}$$

which, from (3.4.2) and $\sum_{j=1}^{q} b_{1j} = 1$, implies

$$\begin{aligned} \tau(\mathcal{E}_n) &= \Big\{ \tau(\boldsymbol{p}) \,\Big|\, L(\boldsymbol{p}) \ge cL(\hat{F}_n),\ \boldsymbol{p} \in \mathcal{F}_n \Big\} \\ &= \Big\{ (\boldsymbol{a}, \boldsymbol{b}) \,\Big|\, 0 \le a_{ij} \le 1,\ 0 < b_{1j} < 1,\ 1 \le j \le q,\ 1 \le i \le m_j; \\ &\qquad\qquad \sum_{j=1}^{q} b_{1j} = 1,\ G(\boldsymbol{a}, \boldsymbol{b}) \ge cL(\hat{F}_n) \Big\}. \end{aligned} \tag{3.4.6}$$

Thus, from (3.4.5) and $G(\boldsymbol{a},\boldsymbol{b}) \geq cL(\hat{F}_n) > 0$, we know that for any $(\boldsymbol{a},\boldsymbol{b}) \in \tau(\mathcal{E}_n)$ we have

$$\begin{cases} 0 < b_{1j} < 1 & \text{for } \ 1 \leq j \leq q \\ 0 \leq a_{ij} < 1 & \text{for } \ (i,j) \in J_1 \\ 0 < a_{ij} \leq 1 & \text{for } \ (i,j) \in J_2 \\ 0 < a_{ij} < 1 & \text{for } \ (i,j) \in J_3, \end{cases} \tag{3.4.7}$$

where J_1, J_2 and J_3 are given by (2.3.4), (2.3.8) and (2.3.12), respectively. If we let

$$\begin{aligned} X \equiv \Big\{ & (\boldsymbol{a},\boldsymbol{b}) \,|\, 0 \leq a_{ij} < 1,\ (i,j) \in J_1;\ 0 < a_{ij} \leq 1,\ (i,j) \in J_2; \\ & 0 < a_{ij} < 1,\ (i,j) \in J_3;\ 0 < b_{1j} < 1,\ 1 \leq j \leq q;\ \sum_{j=1}^{q} b_{1j} = 1 \Big\}, \end{aligned} \tag{3.4.8}$$

then $\tau(\mathcal{E}_n)$ in (3.4.6) is equivalently written as

$$\tau(\mathcal{E}_n) = \Big\{ (\boldsymbol{a},\boldsymbol{b}) \in X \,\Big|\, G(\boldsymbol{a},\boldsymbol{b}) \geq cL(\hat{F}_n) \Big\}. \tag{3.4.9}$$

For any $(\boldsymbol{a}^{(1)},\boldsymbol{b}^{(1)}) \in \tau(\mathcal{E}_n)$ and $(\boldsymbol{a}^{(2)},\boldsymbol{b}^{(2)}) \in \tau(\mathcal{E}_n)$, (3.4.9) implies

$$(\boldsymbol{a}^{(1)},\boldsymbol{b}^{(1)}) \in X, \qquad (\boldsymbol{a}^{(2)},\boldsymbol{b}^{(2)}) \in X \tag{3.4.10}$$

$$G\big(\boldsymbol{a}^{(1)},\boldsymbol{b}^{(1)}\big) \geq cL(\hat{F}_n), \qquad G\big(\boldsymbol{a}^{(2)},\boldsymbol{b}^{(2)}\big) \geq cL(\hat{F}_n). \tag{3.4.11}$$

Since X is a convex set, (3.4.10) implies that for any $0 \leq \lambda \leq 1$,

$$\lambda \begin{pmatrix} \boldsymbol{a}^{(1)} \\ \boldsymbol{b}^{(1)} \end{pmatrix}^{\top} + (1-\lambda) \begin{pmatrix} \boldsymbol{a}^{(2)} \\ \boldsymbol{b}^{(2)} \end{pmatrix}^{\top} \in X. \tag{3.4.12}$$

Also, if we can show

$$G(\boldsymbol{a},\boldsymbol{b}) \quad \text{is} \quad \text{quasiconcave} \quad \text{on} \quad X \tag{3.4.13}$$

(see definition in Bazaraa *et al.*, 1993; page 108), then from (3.4.11) we have

$$\begin{aligned} &G\left(\lambda\begin{pmatrix}\boldsymbol{a}^{(1)}\\ \boldsymbol{b}^{(1)}\end{pmatrix}^{\top} + (1-\lambda)\begin{pmatrix}\boldsymbol{a}^{(2)}\\ \boldsymbol{b}^{(2)}\end{pmatrix}^{\top}\right) \\ &\qquad \geq \min\left\{G\left(\begin{pmatrix}\boldsymbol{a}^{(1)}\\ \boldsymbol{b}^{(1)}\end{pmatrix}^{\top}\right),\ G\left(\begin{pmatrix}\boldsymbol{a}^{(2)}\\ \boldsymbol{b}^{(2)}\end{pmatrix}^{\top}\right)\right\} \geq cL(\hat{F}_n). \end{aligned} \tag{3.4.14}$$

Hence, the convexity of $\tau(\mathcal{E}_n)$ follows from (3.4.9), (3.4.12) and (3.4.14).

To establish (3.4.13), from Bazaraa *et al.* (1993; page 116) it suffices to show that $G(\boldsymbol{a},\boldsymbol{b})$ is concave down on X. Note that $\log G$ is well-defined on X, and from (2.3.31), (2.3.58) and (3.4.8), we have for $(\boldsymbol{a},\boldsymbol{b}) \in X$,

$$\log G(\boldsymbol{a},\boldsymbol{b}) = \log G_1(\boldsymbol{a}_1) + \log G_2(\boldsymbol{a}_2) + \log G_3(\boldsymbol{a}_3) + \log G_5(\boldsymbol{b}), \tag{3.4.15}$$

where $G_1(\boldsymbol{a}_1) : A_1 \to \mathbb{R}$ is given by (2.3.5), $G_2(\boldsymbol{a}_2) : A_2 \to \mathbb{R}$ is given by (2.3.9), $G_3(\boldsymbol{a}_3) : A_3 \to \mathbb{R}$ is given by (2.3.13), $G_5(\boldsymbol{b}) : B \to \mathbb{R}$ is given by (2.3.20), and $A_1,\ A_2,\ A_3$ and B are given by (2.3.51), (2.3.53), (2.3.55) and (2.3.61), respectively. We know that $\log G_3(\boldsymbol{a}_3)$ is concave down on A_3 from (2.3.56), and $\log G_5(\boldsymbol{b})$ is concave down on B from (2.3.66). From (2.3.5), we have for all $\boldsymbol{a}_1 \in A_1$

$$\begin{aligned} &\frac{\partial^2}{\partial a_{ij}^2}\Big(\log G_1(\boldsymbol{a}_1)\Big) = \frac{-N_{ij}}{(1-a_{ij})^2} \ < \ 0 \\ &\frac{\partial^2}{\partial a_{ij}\partial a_{k\ell}}\Big(\log G_1(\boldsymbol{a}_1)\Big) = 0 \qquad \text{for} \quad (i,j) \neq (k,\ell), \end{aligned}$$

which implies $\log G_1(\boldsymbol{a}_1)$ is concave down on A_1. Also, from (2.3.9), we have for all $\boldsymbol{a}_2 \in A_2$

$$\begin{aligned}
&\frac{\partial^2}{\partial a_{ij}^2}\Big(\log G_2(\boldsymbol{a}_2)\Big) = \frac{-\gamma_{ij}}{(a_{ij})^2} \;<\; 0 \\
&\frac{\partial^2}{\partial a_{ij}\partial a_{k\ell}}\Big(\log G_2(\boldsymbol{a}_2)\Big) = 0 \qquad \text{for} \quad (i,j) \neq (k,\ell),
\end{aligned}$$

which implies $\log G_2(\boldsymbol{a}_2)$ is concave down on A_2. Therefore, $\log G$ is concave down on X.

(II) "$\mathcal{E}_n$ is connected": Since $\tau(\mathcal{E}_n)$ is convex, $\tau(\mathcal{E}_n)$ is connected (Royden, 1988; page 183, Problem 35). From (2.3.32), we know $\tau^{-1}(\boldsymbol{a}, \boldsymbol{b})$ exists and is continuous on $\tau(\mathcal{E}_n)$ in (3.4.6). Hence, from Royden (1988; page 182), $\tau^{-1}\big(\tau(\mathcal{E}_n)\big) = \mathcal{E}_n$ is connected. $\square$

Proof of Theorem 3.2: "$\Rightarrow$": Assume $\theta_0 \in S_n = [T_L, T_U]$, where T_L and T_U are given by (3.2.10). From the proof of Theorem 3.1, we know $T(\boldsymbol{p})$ in (3.2.4) is continuous on $\mathcal{E}_n$ in (3.2.9). Thus, since T_L and T_U are the lower bound and upper bound of $T(\boldsymbol{p})$ on $\mathcal{E}_n$, respectively, we know that from the Intermediate Value Theorem, there exists $\boldsymbol{p}^* \in \mathcal{E}_n$ such that $\theta_0 = T(\boldsymbol{p}^*)$. From $\boldsymbol{p}^* \in \mathcal{E}_n$, we know $\boldsymbol{p}^* \in \mathcal{F}_n$ and $L(\boldsymbol{p}^*) \geq cL(\hat{F}_n)$, which implies from $\theta_0 = T(\boldsymbol{p}^*)$ and (3.2.6),

$$r(\theta_0) \;=\; \frac{1}{L(\hat{F}_n)} \sup_{\boldsymbol{p}} \big\{L(\boldsymbol{p}) \,|\, T(\boldsymbol{p}) = \theta_0,\; \boldsymbol{p} \in \mathcal{F}_n\big\} \;\geq\; \frac{L(\boldsymbol{p}^*)}{L(\hat{F}_n)} \;\geq\; c.$$

"$\Leftarrow$": Assume $r(\theta_0) \geq c$, where $r(\theta_0)$ is given by (3.2.6). From (3.2.1), (3.2.4) and (3.2.7), we know that in probability as $n \to \infty$,

$$E_n \equiv \big\{\boldsymbol{p} \,|\, T(\boldsymbol{p}) = \theta_0,\; \boldsymbol{p} \in \mathcal{F}_n\big\} \tag{3.4.16}$$

is not empty. From (3.2.6), we know that for any sufficiently large k, there exists $\boldsymbol{p}^{(k)} \in E_n$

such that

$$\frac{L(\boldsymbol{p}^{(k)})}{L(\hat{F}_n)} \geq r(\theta_0) - \frac{1}{k},$$

which, from $r(\theta_0) \geq c$, implies

$$\boldsymbol{p}^{(k)} \in \mathcal{F}_n, \qquad T(\boldsymbol{p}^{(k)}) = \theta_0, \qquad \frac{L(\boldsymbol{p}^{(k)})}{L(\hat{F}_n)} \geq c - \frac{1}{k}. \tag{3.4.17}$$

Since $\boldsymbol{p}^{(k)} \in \mathcal{F}_n$ and $\mathcal{F}_n$ is compact, we know that $\{\boldsymbol{p}^{(k)}\}$ is bounded, thus there exists a convergent subsequence, still denoted as $\boldsymbol{p}^{(k)}$, such that

$$\boldsymbol{p}^{(k)} \to \boldsymbol{p}^{(0)} \in \mathcal{F}_n, \qquad \text{as} \;\; k \to \infty. \tag{3.4.18}$$

Since $L(\boldsymbol{p})$ is continuous, from (3.4.17)-(3.4.18) we have

$$\frac{L(\boldsymbol{p}^{(0)})}{L(\hat{F}_n)} = \lim_{k\to\infty} \frac{L(\boldsymbol{p}^{(k)})}{L(\hat{F}_n)} \geq c > 0, \tag{3.4.19}$$

which, from the arguments in (3.4.1)-(3.4.3), implies $\sum_{k=1}^{m_\zeta+1} p_{k\zeta}^{(0)} > 0$; in turn, from (3.4.17)-(3.4.18) and (3.2.4) we have

$$T(\boldsymbol{p}^{(0)}) = \lim_{k\to\infty} T(\boldsymbol{p}^{(k)}) = \theta_0. \tag{3.4.20}$$

The proof follows from (3.4.18)-(3.4.20) and Theorem 3.1. $\square$

3.5 Proof of Theorem 3.3

Before proving Theorem 3.3, we establish the following lemmas, while the proofs are given at the end of this section.

Lemma 3.1. *For Λ_n and $g(\lambda)$ given by (AS3.1) and (3.2.14), respectively, we have as*

$n \to \infty$,

$$
\begin{aligned}
&(i) \qquad P\{V \le t_0, Z = z_0, \delta = 1\} > 0, \qquad \textit{in probability} \\
&(ii) \qquad E_{n1} \equiv \big\{ i \,|\, 1 \le i \le m_\zeta,\ U_i \le t_0,\ \gamma_{i\zeta} > 0 \big\} \ne \emptyset, \qquad \textit{in probability} \\
&(iii) \qquad \exists\, a \ \textit{unique solution to}\ g(\lambda) = 0 \ \textit{on}\ (0, \Lambda_n), \qquad \textit{in probability.}
\end{aligned}
$$

Lemma 3.2. *Let* $\tilde{\lambda}$ *denote the unique solution to* $g(\lambda) = 0$ *on* $(0, \Lambda_n)$ *in Lemma 3.1 and let*

$$
\mathcal{J}_1 \equiv \big\{ (i, \zeta) \,|\, 1 \le i \le m_\zeta,\ U_i \le t_0 \big\} \tag{3.5.1}
$$

$$
G_6(\boldsymbol{a}_6) \equiv \prod_{(i,\zeta) \in \mathcal{J}_1} (a_{i\zeta})^{\gamma_{i\zeta}} \big(1 - a_{i\zeta}\big)^{N_{i\zeta} - \gamma_{i\zeta}}, \qquad \boldsymbol{a}_6 = \{ a_{i\zeta} \,|\, (i, \zeta) \in \mathcal{J}_1 \}. \tag{3.5.2}
$$

Then, the solution to

$$
\begin{cases}
\max\ G_6(\boldsymbol{a}_6) \\
\text{subject to:} \quad 0 \le a_{i\zeta} \le 1, \quad (i, \zeta) \in \mathcal{J}_1; \quad 1 - \theta_0 = \displaystyle\prod_{(i,\zeta) \in \mathcal{J}_1} \big(1 - a_{i\zeta}\big)
\end{cases} \tag{3.5.3}
$$

is uniquely given by $\tilde{\boldsymbol{a}}_6 = \big\{ \tilde{a}_{i\zeta} \,|\, (i, \zeta) \in \mathcal{J}_1 \big\}$, *where*

$$
\tilde{a}_{i\zeta} = \frac{\gamma_{i\zeta}}{N_{i\zeta} - \tilde{\lambda}}, \qquad (i, \zeta) \in \mathcal{J}_1. \tag{3.5.4}
$$

Lemma 3.3. *Let*

$$
\mathcal{J}_2 \equiv \big\{ (i, \zeta) \,|\, 1 \le i \le m_\zeta,\ U_i > t_0 \big\} \cup \big\{ (i, j) \,|\, j \ne \zeta,\ 1 \le i \le m_j \big\}. \tag{3.5.5}
$$

For $G_4(\boldsymbol{a})$, J *and* $\mathcal{J}_1$ *given by (2.3.17), (2.3.16) and (3.5.1), respectively, we have* $J = \mathcal{J}_1 \cup \mathcal{J}_2$

and the solution to

$$\begin{cases} \max\ G_4(\boldsymbol{a}) \\ \text{subject to}: \quad 0 \le a_{ij} \le 1, \quad (i,j) \in J; \;\; 1-\theta_0 = \displaystyle\prod_{(i,\zeta)\in\mathcal{J}_1} \big(1-a_{i\zeta}\big) \end{cases} \tag{3.5.6}$$

is uniquely given by $\tilde{\boldsymbol{a}} = \big\{\tilde{a}_{ij} \,|\, (i,j) \in J\big\}$, *where*

$$\tilde{a}_{ij} = \begin{cases} \dfrac{\gamma_{ij}}{N_{ij}-\tilde{\lambda}}, & (i,j) \in \mathcal{J}_1 \\ \dfrac{\gamma_{ij}}{N_{ij}}, & (i,j) \in \mathcal{J}_2. \end{cases} \tag{3.5.7}$$

Proof of Theorem 3.3: To obtain an expression for $r(\theta_0)$ in (3.2.6), we solve the following optimization problem

$$\begin{cases} \max\ \ L(\boldsymbol{p}) \\ \text{subject to}: \quad 0 \le p_{ij} \le 1, \;\; 1 \le j \le q, \;\; 1 \le i \le m_j; \\ \qquad\qquad\qquad \sum_{j=1}^{q} \sum_{i=1}^{m_j+1} p_{ij} = 1; \quad \theta_0 = T(\boldsymbol{p}). \end{cases} \tag{3.5.8}$$

Under transformation (2.3.23), we know from (2.3.28) and (3.3.3) that optimization problem (3.5.8) is equivalent to

$$\begin{cases} \max\ G(\boldsymbol{a},\boldsymbol{b}) \\ \text{subject to}: \quad 0 \le a_{ij} \le 1, \;\; 0 \le b_{1j} \le 1, \;\; (i,j) \in J; \\ \qquad\qquad\qquad \sum_{j=1}^{q} b_{1j} = 1; \quad 1-\theta_0 = T_1(\boldsymbol{a}), \end{cases} \tag{3.5.9}$$

where J is given by (2.3.16) and $\boldsymbol{a}$ and $\boldsymbol{b}$ are given by (2.3.29). Note that $T_1(\boldsymbol{a})$ in (3.3.3) is a polynomial, thus it is continuous in $\boldsymbol{a}$. Therefore, the constraint set

$$\Big\{(\boldsymbol{a},\boldsymbol{b}) \,\Big|\, 0 \le a_{ij} \le 1, \;\; 0 \le b_{1j} \le 1, \;\; (i,j) \in J; \;\; \sum_{j=1}^{q} b_{1j} = 1, \;\; 1-\theta_0 = T_1(\boldsymbol{a})\Big\}$$

is compact. Since $G(\boldsymbol{a},\boldsymbol{b})$ in (2.3.28) is a polynomial, it is continuous in $(\boldsymbol{a},\boldsymbol{b})$, thus we know a solution to (3.5.9) exists. From (3.3.3) and (3.5.1), we know that the constraint set can be written as

$$
\begin{aligned}
&\Big\{(\boldsymbol{a},\boldsymbol{b})\Big|\, 0\le a_{ij}\le 1,\ 0\le b_{1j}\le 1,\ (i,j)\in J;\ \sum_{j=1}^{q} b_{1j}=1,\ 1-\theta_0=\prod_{(i,\zeta)\in\mathcal{J}_1}\big(1-a_{i\zeta}\big)\Big\}\\
&=\Big\{\boldsymbol{a}\,\Big|\, 0\le a_{ij}\le 1,\ (i,j)\in J,\ 1-\theta_0=\prod_{(i,\zeta)\in\mathcal{J}_1}\big(1-a_{i\zeta}\big)\Big\}\\
&\qquad\qquad \cup\Big\{\boldsymbol{b}\,\Big|\, 0\le b_{1j}\le 1,\ 1\le j\le q;\ \sum_{j=1}^{q} b_{1j}=1\Big\},
\end{aligned}
$$

and since $G(\boldsymbol{a},\boldsymbol{b})$ in (2.3.28) can be written as $G(\boldsymbol{a},\boldsymbol{b})=G_4(\boldsymbol{a})\,G_5(\boldsymbol{b})$, where $G_4(\boldsymbol{a})$ and $G_5(\boldsymbol{b})$ are given by (2.3.17) and (2.3.20), respectively, we know from Lemmas 3.3 and 2.5 that the solution for (3.5.9) is uniquely given by $(\tilde{\boldsymbol{a}},\hat{\boldsymbol{b}})$, where $\tilde{\boldsymbol{a}}$ is the unique solution to optimization problem (3.5.6) and is given by (3.5.7) and $\hat{\boldsymbol{b}}$ is the unique solution to optimization problem (2.3.21) and is given by (2.3.22).

Therefore, from (3.5.8)-(3.5.9), we have the following expression for $r(\theta_0)$ in (3.2.6):

$$
r(\theta_0)=\frac{G(\tilde{\boldsymbol{a}},\hat{\boldsymbol{b}})}{L(\hat{F}_n)}, \tag{3.5.10}
$$

where $\tilde{\boldsymbol{a}}$ is given by (3.5.7), $\hat{\boldsymbol{b}}$ is given by (2.3.22), and we know from (2.3.28) and (2.3.30) that $L(\hat{F}_n)=G(\hat{\boldsymbol{a}},\hat{\boldsymbol{b}})$ with $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$ given by (2.3.19) and (2.3.22), respectively. Noting that $\hat{a}_{ij}=\tilde{a}_{ij}$ for $(i,j)\in\mathcal{J}_2$, we have from (2.3.28), (3.5.1) and (3.5.10)

$$
\begin{aligned}
r(\theta_0)=\frac{G(\tilde{\boldsymbol{a}},\hat{\boldsymbol{b}})}{G(\hat{\boldsymbol{a}},\hat{\boldsymbol{b}})}
&=\frac{\left(\prod_{j=1}^{q}(\hat{b}_{1j})^{N_{1j}}\right)\prod\prod_{(i,j)\in J}(\tilde{a}_{ij})^{\gamma_{ij}}(1-\tilde{a}_{ij})^{N_{ij}-\gamma_{ij}}}{\left(\prod_{j=1}^{q}(\hat{b}_{1j})^{N_{1j}}\right)\prod\prod_{(i,j)\in J}(\hat{a}_{ij})^{\gamma_{ij}}(1-\hat{a}_{ij})^{N_{ij}-\gamma_{ij}}}\\
&=\prod\prod_{(i,j)\in\mathcal{J}_1}\left(\frac{\tilde{a}_{ij}}{\hat{a}_{ij}}\right)^{\gamma_{ij}}\left(\frac{1-\tilde{a}_{ij}}{1-\hat{a}_{ij}}\right)^{N_{ij}-\gamma_{ij}}=\prod_{\substack{i=1\\ U_i\le t_0}}^{m_\zeta}\left(\frac{\tilde{a}_{ij}}{\hat{a}_{ij}}\right)^{\gamma_{ij}}\left(\frac{1-\tilde{a}_{ij}}{1-\hat{a}_{ij}}\right)^{N_{ij}-\gamma_{ij}}.
\end{aligned}
$$

Plugging in $\tilde{a}_{ij}$ and $\hat{a}_{ij}$ in (3.5.7) and (2.3.19), respectively, and doing some simple algebra, we obtain the expression for $\log r(\theta_0)$ in (3.2.13). $\square$

Proof of Lemma 3.1 (i): Note that as $n \to \infty$, we have

$$U_1 < t_0, \qquad \text{in probability,} \tag{3.5.11}$$

where the proof of (3.5.11) is similar to that of (3.2.1). From (3.5.11), we know that as $n \to \infty$, there exists $\xi_0 > 0$ and $t_1 \le t_0$, where t_1 is an interior point of the support of C, such that $\bar{F}_C(t) \ge c_0$ for all $t_1 \le t \le t_0$. Therefore, we have

$$\begin{aligned}
&P\{V \le t_0, Z = z_0, \delta = 1\} = P\{T \le t_0, Z = z_0, T \le C\} \\
&\qquad = \iint\limits_{\substack{t \le t_0 \\ t \le c}} f(t, c, z_0)\, dt\, dc \;=\; \iint\limits_{\substack{t \le t_0 \\ t \le c}} f_0(t, z_0) f_C(c)\, dt\, dc \\
&\qquad = \int_{-\infty}^{t_0} f_0(t, z_0) \int_t^{\infty} f_C(c)\, dc\, dt \;=\; \int_{-\infty}^{t_0} f_0(t, z_0) \bar{F}_C(t)\, dt \\
&\qquad \ge \int_{t_1}^{t_0} f_0(t, z_0) \bar{F}_C(t)\, dt \;\ge\; \xi_0 \int_{t_1}^{t_0} f_0(t, z_0)\, dt \;>\; 0, \qquad \text{in probability.} \qquad \square
\end{aligned}$$

Proof of Lemma 3.1 (ii): From (2.2.19), we know $N_{k\zeta} > 0$ for all $1 \le k \le m_\zeta$, which implies that E_{n1} in the statement of Lemma 3.1 (ii) can be written as

$$E_{n1} = \big\{k \,|\, 1 \le k \le m_\zeta,\ U_k \le t_0,\ \gamma_{k\zeta} > 0,\ N_{k\zeta} > 0\big\}. \tag{3.5.12}$$

From $N_{k\zeta} > 0$ and (2.2.16), we have for all $1 \le k \le m_\zeta$

$$\exists\ (V_i, Z_i) \ \text{ such that } \ V_i \ge U_k \ \text{ and } \ Z_i = W_\zeta = z_0, \tag{3.5.13}$$

which implies

$$\begin{aligned}
E_{n1} \neq \emptyset \quad &\Longleftrightarrow \quad \exists\, 1 \le k \le m_\zeta \;\text{ s.t. }\; U_k \le t_0,\; \gamma_{k\zeta} > 0,\; N_{k\zeta} > 0 \\
&\Longleftrightarrow \quad \exists\,(V_i, Z_i) \text{ and } \exists\, 1 \le k \le m_\zeta \;\text{ s.t. }\; V_i \ge U_k,\; U_k \le t_0,\; Z_i = z_0,\; \gamma_{k\zeta} > 0,
\end{aligned}$$

in turn, we know that E_{n1} in (3.5.12) can be written as

$$E_{n1} = \big\{k \,|\, \exists\,(V_i, Z_i) \text{ satisfying } V_i \ge U_k,\, U_k \le t_0,\, Z_i = z_0,\, \gamma_{k\zeta} > 0,\, 1 \le k \le m_\zeta\big\}. \quad (3.5.14)$$

If we let

$$A_n \equiv \{V_i \,|\, V_i \le t_0,\; Z_i = z_0,\; 1 \le i \le n\}, \quad (3.5.15)$$

then from (3.5.11) and (3.2.2), we have that as $n \to \infty$, there exist $U_{\alpha_1}, \cdots, U_{\alpha_\rho}$, in probability, which denote the distinct V_i's in A_n such that $U_{\alpha_k} \le t_0$ for $1 \le k \le \rho$. Therefore, we have from (2.2.4)

$$\begin{aligned}
\gamma_{\alpha_k,\zeta} &= \sum_{i=1}^{n} I\{V_i = U_{\alpha_k},\, \delta_i = 1,\; Z_i = z_0\} \\
&= [\#\text{ of } (V_i, Z_i, \delta_i)'\text{s } \text{ satisfying } V_i = U_{\alpha_k} \le t_0,\; Z_i = z_0,\; \delta_i = 1]. \quad (3.5.16)
\end{aligned}$$

and we see that if $\gamma_{\alpha_k,\zeta} > 0$ for some $1 \le k \le \rho$, then $\alpha_k \in E_{n1}$, which implies

$$P\{E_{n1} \neq \emptyset\} \ge P\{\exists\, \gamma_{\alpha_k\zeta} > 0 \text{ for some } 1 \le k \le \rho\}. \quad (3.5.17)$$

Note that from Lemma 3.1 (i), we have

$$\begin{aligned}
P\{\exists\, \gamma_{\alpha_k\zeta} > 0 \text{ for some } 1 \le k \le \rho\} &= 1 - P\{\gamma_{\alpha_k\zeta} = 0 \text{ for all } 1 \le k \le \rho\} \\
&= 1 - P\big\{[\#\text{ of } (V_i, Z_i, \delta_i)'\text{s s.t. } V_i = U_{\alpha_k} \le t_0,\; Z_i = z_0,\; \delta_i = 1] = 0,\; 1 \le k \le \rho\big\}
\end{aligned}$$

$$= 1 - P\big\{[\#\text{ of } (V_i, Z_i, \delta_i)'\text{s s.t. } V_i \le t_0,\ Z_i = z_0,\ \delta_i = 1] = 0\big\}$$

$$= 1 - P\{(V_i, Z_i, \delta_i) \in A^c,\ 1 \le i \le n\} \ = \ 1 - \prod_{i=1}^{n} P\{(V_i, Z_i, \delta_i) \in A^c\}$$

$$= 1 - \big[P\{(V, Z, \delta) \in A^c\}\big]^n \ = \ 1 - \big[1 - P\{(V, Z, \delta) \in A\}\big]^n$$

$$= 1 - \big[1 - P\{V \le t_0,\, Z = z_0,\, \delta = 1\}\big]^n \to 1 \qquad \text{as } \ n \to \infty, \tag{3.5.18}$$

where $A \equiv \{(t, z, d) \,|\, t \le t_0,\ z = z_0,\ d = 1\}$. The proof follows from (3.5.17)-(3.5.18). □

Proof of Lemma 3.1 (iii): Note that as $n \to \infty$, we know

$$0 < \theta_0 < 1, \qquad \text{in probability}, \tag{3.5.19}$$

which implies that $g(\lambda)$ in (3.2.14) is well-defined on $(0, \Lambda_n)$, in probability. Therefore, $g(\lambda)$ is continuous on $(0, \Lambda_n)$, in probability, since it is a log function and is well-defined on $(0, \Lambda_n)$, in probability. As $n \to \infty$, we know $g(\lambda)$ is a strictly increasing function on $(0, \Lambda_n)$, in probability, since from Lemma 3.1 (ii) and (3.2.14), we have in probability for $\lambda \in (0, \Lambda_n)$,

$$g'(\lambda) = \sum_{\substack{i=1 \\ U_i \le t_0}}^{m_\zeta} \frac{1}{1 - \frac{\gamma_{i\zeta}}{N_{i\zeta} - \lambda}} \Big(\frac{\gamma_{i\zeta}}{(N_{i\zeta} - \lambda)^2} \Big) = \sum_{\substack{i=1 \\ U_i \le t_0}}^{m_\zeta} \frac{\gamma_{i\zeta}}{(N_{i\zeta} - \lambda)(N_{i\zeta} - \lambda - \gamma_{i\zeta})} > 0. \tag{3.5.20}$$

In addition, note that from (AS3.2), we have as $n \to \infty$

$$\lim_{\lambda \to 0+} g(\lambda) = \log(1 - \theta_0) - \sum_{\substack{i=1 \\ U_i \le t_0}}^{m_\zeta} \log\Big(1 - \frac{\gamma_{i\zeta}}{N_{i\zeta}}\Big) < 0, \qquad \text{in probability} \tag{3.5.21}$$

$$\lim_{\lambda \to \Lambda_n -} g(\lambda) = \log(1 - \theta_0) - \sum_{\substack{i=1 \\ U_i \le t_0}}^{m_\zeta} \log\Big(1 - \frac{\gamma_{i\zeta}}{N_{i\zeta} - \Lambda_n}\Big) > 0, \qquad \text{in probability.} \tag{3.5.22}$$

The proof follows from the Intermediate Value Theorem and (3.5.20)-(3.5.22). □

Proof of Lemma 3.2: From (3.5.19), we know that as $n \to \infty$, any solution to (3.5.3) will satisfy the following, in probability,

$$\begin{cases} 0 \le a_{ij} < 1 & \text{if} \quad \gamma_{ij} = 0, \quad (i,j) \in \mathcal{J}_1 \\ 0 < a_{ij} < 1 & \text{if} \quad \gamma_{ij} > 0, \quad (i,j) \in \mathcal{J}_1. \end{cases} \tag{3.5.23}$$

If we let

$$h(\boldsymbol{a}_6) \equiv \log(1-\theta_0) - \sum_{(i,\zeta)\in\mathcal{J}_1} \log\big(1-a_{i\zeta}\big) \tag{3.5.24}$$

$$\mathcal{A} \equiv \big\{\boldsymbol{a}_6 \,|\, 0 < a_{ij} < 1,\ (i,j) \in \mathcal{J}_1\big\}, \tag{3.5.25}$$

then to solve optimization problem (3.5.3), we solve the following optimization problem

$$\begin{cases} \min\ -\log G_6(\boldsymbol{a}_6) = \displaystyle\sum_{(i,\zeta)\in\mathcal{J}_1} \Big[-\gamma_{i\zeta}\log a_{i\zeta} - (N_{i\zeta}-\gamma_{i\zeta})\log(1-a_{i\zeta})\Big] \\ \text{subject to :} \quad \boldsymbol{a}_6 \in \mathcal{A}, \quad h(\boldsymbol{a}_6) = 0. \end{cases} \tag{3.5.26}$$

As follows, we discuss the KKT sufficient conditions (Bazaraa *et al.*, 1993; page 164) for optimization problem (3.5.26).

Using Lagrange multipliers, let

$$\begin{aligned} \mathcal{G}(\boldsymbol{a}_6,\lambda) &\equiv -\log G_6(\boldsymbol{a}_6) + \lambda h(\boldsymbol{a}_6) \\ &= \sum_{(i,\zeta)\in\mathcal{J}_1} \Big[-\gamma_{i\zeta}\log a_{i\zeta} - (N_{i\zeta}-\gamma_{i\zeta}+\lambda)\log(1-a_{i\zeta})\Big] + \lambda\log(1-\theta_0), \end{aligned} \tag{3.5.27}$$

which has the following partial derivatives

$$\begin{cases} \dfrac{\partial}{\partial a_{i\zeta}}\Big(\mathcal{G}(\boldsymbol{a}_6,\lambda)\Big) = \dfrac{-\gamma_{i\zeta}}{a_{i\zeta}} + \dfrac{N_{i\zeta}-\gamma_{i\zeta}+\lambda}{1-a_{i\zeta}}, \qquad (i,\zeta)\in\mathcal{J}_1 \\ \dfrac{\partial}{\partial\lambda}\Big(\mathcal{G}(\boldsymbol{a}_6,\lambda)\Big) = \log(1-\theta_0) - \displaystyle\sum_{(i,\zeta)\in\mathcal{J}_1}\log\big(1-a_{i\zeta}\big). \end{cases} \tag{3.5.28}$$

From $\partial\mathcal{G}/\partial a_{i\zeta} = 0$, we have

$$a_{ij\zeta} = \frac{\gamma_{i\zeta}}{N_{i\zeta} - \lambda}, \qquad (i,\zeta) \in \mathcal{J}_1, \tag{3.5.29}$$

and from $\partial\mathcal{G}/\partial\lambda = 0$, (3.5.29) and (3.5.1), we have

$$\log(1-\theta_0) - \sum_{\substack{i=1\\U_i\le t_0}}^{m_\zeta} \log\Big(1 - \frac{\gamma_{i\zeta}}{N_{i\zeta} - \lambda}\Big) = 0. \tag{3.5.30}$$

Therefore, we see that $\tilde{\boldsymbol{a}}_6 = \big\{\tilde{a}_{ij} \,|\, (i,j) \in \mathcal{J}_1\big\}$ given by (3.5.4) is a solution to $\nabla\mathcal{G}(\boldsymbol{a}_6, \lambda) = -\nabla \log G_6(\boldsymbol{a}_6) + \lambda\nabla h(\boldsymbol{a}_6) = \boldsymbol{0}$ since $\tilde{\lambda}$ is a solution to $g(\lambda) = 0$, where $g(\lambda)$ is given by (3.2.14).

Note that $\mathcal{A}$ in (3.5.25) is open and convex. Also, note that we can relax the restriction on $\mathcal{A}$ to (3.5.23) because if $\gamma_{ij} = 0$, then $\tilde{a}_{ij} = 0$ and we have in the sum for $\log G_6(\boldsymbol{a}_6)$ in (3.5.2)

$$\gamma_{ij} \log a_{ij} = 0 \log 0 = 0,$$

thus $\log G_6(\boldsymbol{a}_6)$ is well-defined. Noting that $\gamma_{i\zeta} \ge 0$ for all $1 \le i \le m_\zeta$, we see that $\tilde{a}_{ij}$ in (3.5.4) satisfies (3.5.23) since $\tilde{\lambda} \in (0, \Lambda_n)$. Note that $-\log G_6(\boldsymbol{a}_6)$ in (3.5.2) is strictly concave up on $\mathcal{A}$ because

$$\begin{cases} \dfrac{\partial^2}{\partial a_{ij}^2}\Big(-\log G_6(\boldsymbol{a}_6)\Big) = \dfrac{\gamma_{ij}}{(a_{ij})^2} + \dfrac{N_{ij}-\gamma_{ij}}{(1-a_{ij}^*)^2} \;>\; 0 \\ \dfrac{\partial^2}{\partial a_{ij}\partial a_{k\ell}}\Big(-\log G_6(\boldsymbol{a}_6)\Big) = 0 \quad \text{for} \quad (i,j) \ne (k,\ell). \end{cases}$$

Therefore, from Bazaraa *et al.* (1993; page 116), we know $-\log G_6(\boldsymbol{a}_6)$ is pseudoconvex on $\mathcal{A}$ (see definition in Bazaraa *et al.*, 1993; page 113). Also, from Bazaraa *et al.* (1993; page 116) we know $h(\boldsymbol{a}_6)$ in (3.5.24) is quasiconvex on $\mathcal{A}$ (see definition in Bazaraa *et al.*, 1993;

page 108) since

$$\begin{cases} \dfrac{\partial^2}{\partial a_{ij}^2}\Big(h(\boldsymbol{a}_6)\Big) = \dfrac{1}{(1-a_{ij})^2} \; > \; 0 \\ \dfrac{\partial^2}{\partial a_{ij}\partial a_{k\ell}}\Big(h(\boldsymbol{a}_6)\Big) = 0 \quad \text{for} \quad (i,j) \neq (k,\ell), \end{cases}$$

which implies that $h(\boldsymbol{a}_6)$ is strictly concave up on $\mathcal{A}$. Since $\tilde{\lambda} > 0$ is the unique solution to $g(\lambda) = 0$, the proof follows from the KKT sufficient conditions and Theorem 3.4.2 in Bazaraa *et al.* (1993; pages 164 and 101, respectively). $\square$

Proof of Lemma 3.3: Note that from (2.3.13), (3.5.1) and (3.5.5), we have $\mathcal{J}_1 \cap \mathcal{J}_2 = \emptyset$ and $J = \mathcal{J}_1 \cup \mathcal{J}_2$. Therefore, the constraint set in (3.5.6) can be written as

$$\begin{aligned} &\Big\{\boldsymbol{a} \,\Big|\, 0 \leq a_{ij} \leq 1,\ (i,j) \in J;\ 1-\theta_0 = \prod_{(i,\zeta)\in\mathcal{J}_1} \big(1-a_{i\zeta}\big)\Big\} \\ &= \Big\{\boldsymbol{a} \,\Big|\, 0 \leq a_{ij} \leq 1,\ (i,j) \in \mathcal{J}_1,\ 1-\theta_0 = \prod_{(i,\zeta)\in\mathcal{J}_1} \big(1-a_{i\zeta}\big)\Big\} \\ &\qquad \cup \Big\{\boldsymbol{a} \,\big|\, 0 \leq a_{ij} \leq 1,\ (i,j) \in \mathcal{J}_2\Big\}, \end{aligned}$$

and $G_4(\boldsymbol{a})$ in (2.3.14) can be written as $G_4(\boldsymbol{a}) = G_6(\boldsymbol{a}_6)\, G_7(\boldsymbol{a}_7)$, where $G_6(\boldsymbol{a}_6)$ is given by (3.5.2) and

$$G_7(\boldsymbol{a}_7) \equiv \prod_{(i,j)\in\mathcal{J}_2} (a_{ij})^{\gamma_{ij}} \big(1-a_{ij}\big)^{N_{ij}-\gamma_{ij}}, \qquad \boldsymbol{a}_7 = \{a_{ij} \,|\, (i,j) \in \mathcal{J}_2\}. \tag{3.5.31}$$

From Lemma 2.4, we know the solution to

$$\begin{cases} \max\, G_7(\boldsymbol{a}_7) = \displaystyle\prod\prod_{(i,j)\in\mathcal{J}_2} (a_{ij})^{\gamma_{ij}} (1-a_{ij})^{N_{ij}-\gamma_{ij}} \\ \text{subject to}: \quad 0 \leq a_{ij} \leq 1,\ (i,j) \in \mathcal{J}_2 \end{cases} \tag{3.5.32}$$

is uniquely given by $\tilde{\boldsymbol{a}}_7 = \left\{\tilde{a}_{ij} \,|\, (i,j) \in \mathcal{J}_2\right\}$, where

$$\tilde{a}_{ij} = \frac{\gamma_{ij}}{N_{ij}}, \qquad (i,j) \in \mathcal{J}_2 \tag{3.5.33}$$

because the constraint set for optimization problem (3.5.32) is a subset of the constraint set for optimization problem (2.3.18) and functions G_4 and G_7 have a similar form. Therefore, from $J = \mathcal{J}_1 \cup \mathcal{J}_2$, Lemma 3.2 and (3.5.32)-(3.5.33), we know that the solution for (3.5.6) is uniquely given by

$$\tilde{\boldsymbol{a}} = \left\{\tilde{a}_{ij} \,|\, (i,j) \in J\right\} = \tilde{\boldsymbol{a}}_6 \cup \tilde{\boldsymbol{a}}_7, \tag{3.5.34}$$

where $\tilde{\boldsymbol{a}}_6$ is the unique solution to optimization problem (3.5.3) and is given by (3.5.4) and $\tilde{\boldsymbol{a}}_7$ is uniquely given by (3.5.33). $\square$

CHAPTER 4. WEIGHTED EMPIRICAL LIKELIHOOD-BASED MAXIMUM LIKELIHOOD ESTIMATOR FOR COX MODEL

In this chapter, we derive the weighted empirical likelihood-based estimators for Cox model (1.2.24) in a unified form for various types of censored data mentioned in Section 1.3.

4.1 Introduction

Let

$$(\boldsymbol{O}_i, Z_i), \qquad 1 \leq i \leq n \tag{4.1.1}$$

be the observed data on sample (1.2.17), where $\boldsymbol{O}_i$'s are the observed censored data on T_i's, and the censoring can be right censoring (1.3.2), doubly censoring (1.3.6), interval censoring (1.3.8)-(1.3.9), or partly-interval censoring (1.3.12)-(1.3.13), etc. We denote

$$\hat{F}_n(t, z) = \sum_{i=1}^{m} \sum_{j=1}^{q} \hat{\omega}_{ij} I\{U_i \leq t, W_j \leq q\} \tag{4.1.2}$$

as an estimator for joint d.f. $F_0(t, z)$ for (T, Z) based on observed censored data (4.1.1), where $\hat{\omega}_{ij}$ is the probability mass at point (U_i, W_j) computed based on (4.1.1) satisfying $U_1 < \cdots < U_m$ and $W_1 < \cdots < W_q$. In the case of right censored data (1.2.23) or (2.1.2), we have $\boldsymbol{O}_i = (V_i, \delta_i)$ in (4.1.1) and $\hat{\omega}_{ij} = \hat{q}_{ij}$ in (4.1.2) where $\hat{q}_{ij}$'s are given by (2.2.18). For doubly censored data, see Ren and Gu (1997). We refer to Ren and He (2011) for discussions on other types of censored data. In this chapter, we derive weighted empirical likelihood-based estimator $\hat{\beta}$ for β_0 in Cox model (1.2.24) with censored data (4.1.1).

In Section 4.2, we derive the weighted empirical likelihood function (Ren, 2001) for Cox model (1.2.24) with censored data (4.1.1). Then, the weighted empirical likelihood-based estimator $\hat{\beta}$ for β_0 is derived in Section 4.3. Some remarks on the weighted empirical likelihood function and estimator $\hat{\beta}$ are given in Section 4.4.

4.2 Weighted Empirical Likelihood Function for Cox Model

For Cox model (1.2.24), let $G_0(t)$ and $g_0(t)$ denote the d.f. and the probability density function (p.d.f.), respectively, corresponding to baseline hazard function $h_0(t)$ and let $F(t|z)$ and $f(t|z)$ denote the conditional d.f. and conditional p.d.f., respectively, of T given $Z = z$. Then, under Cox model assumption (1.2.24), we have from (1.2.4)

$$\begin{aligned} \bar{F}(t|z) &= \exp\Big\{ -\int_0^t h(u;z)du\Big\} = \exp\Big\{ -e^{\beta_0 z}\int_0^t h_0(u)du\Big\} \\ &= \exp\big\{ -e^{\beta_0 z}H_0(t)\big\} = \big(\exp\big\{ -H_0(t)\big\}\big)^{e^{\beta_0 z}} = \big(\bar{G}_0(t)\big)^{e^{\beta_0 z}} \\ \Longleftrightarrow \qquad f(t|z) &= e^{\beta_0 z} g_0(t)\big(\bar{G}_0(t)\big)^{e^{\beta_0 z}-1}. \end{aligned} \tag{4.2.1}$$

If we let $f_0(t,z)$ and $f_z(z)$ denote the p.d.f. of $F_0(t,z)$ and the marginal p.d.f. of Z, respectively, then (4.2.1) implies

$$f_0(t,z) = f(t|z)f_z(z) = e^{\beta_0 z} f_z(z) g_0(t)\big(\bar{G}_0(t)\big)^{e^{\beta_0 z}-1}. \tag{4.2.2}$$

Applying (4.1.2) to (1.4.25), we obtain the bivariate version of the weighted empirical likelihood function (Ren, 2001) under Cox model (1.2.24) with censored data (4.1.1) as follows:

$$\begin{aligned} \prod_{i=1}^m \prod_{j=1}^q \Big[f_0(U_i,W_j)\Big]^{n\hat{\omega}_{ij}} &= \prod_{i=1}^m \prod_{j=1}^q \Big[e^{\beta_0 W_j} f_z(W_j) g_0(U_i)\big(\bar{G}_0(U_i)\big)^{e^{\beta_0 W_j}-1}\Big]^{n\hat{\omega}_{ij}} \\ &= \Big(\prod_{i=1}^m \prod_{j=1}^q \Big[f_z(W_j)\Big]^{n\hat{\omega}_{ij}}\Big)\Big(\prod_{i=1}^m \prod_{j=1}^q e^{\beta_0 W_j}\Big[g_0(U_i)\Big]^{n\hat{\omega}_{ij}}\Big[\bar{G}_0(U_i)\Big]^{n\hat{\omega}_{ij}\left(e^{\beta_0 W_j}-1\right)}\Big) \end{aligned}$$

$$= \left(\prod_{i=1}^{m} \prod_{j=1}^{q} \Big[f_z(W_j) \Big]^{n\hat{\omega}_{ij}} \right)$$
$$\times \left(\prod_{i=1}^{m} e^{\beta_0 \sum_{j=1}^{q} W_j} \Big[g_0(U_i) \Big]^{n \sum_{j=1}^{q} \hat{\omega}_{ij}} \Big[\bar{G}_0(U_i) \Big]^{n \sum_{j=1}^{q} \hat{\omega}_{ij} \left(e^{\beta_0 W_j} - 1 \right)} \right). \tag{4.2.3}$$

Thus, the weighted empirical likelihood function for (β_0, G_0) under Cox model (1.2.24) based on data (4.1.1) is proportional to

$$L(\beta, G) = \prod_{i=1}^{m} e^{\beta \mathbb{W}} \Big[G(U_i) - G(U_i -) \Big]^{n\hat{\omega}_i} \Big[\bar{G}(U_i) \Big]^{n\omega_i(\beta)}, \tag{4.2.4}$$

where G is any distribution function, and

$$\mathbb{W} \equiv \sum_{j=1}^{q} W_j, \qquad \hat{\omega}_i \equiv \sum_{j=1}^{q} \hat{\omega}_{ij}, \qquad \text{and} \quad \omega_i(\beta) \equiv \sum_{j=1}^{q} \hat{\omega}_{ij} \big(e^{\beta W_j} - 1 \big). \tag{4.2.5}$$

In the next section, we derive the weighted empirical likelihood-based estimator $(\hat{\beta}, \hat{G})$ for (β_0, G_0) under likelihood function (4.2.4).

4.3 Weighted Empirical Likelihood-Based Maximum Likelihood Estimator

To maximize (4.2.4), we consider the case where $\beta > 0$ and without loss of generality, assume $W_j > 0$ for all $1 \le j \le q$, because otherwise we can just shift the W_j's. We restrict all possible candidates to those d.f.'s that assign all their probability masses to points (U_i, W_j) and line segment $L = \{t \in \mathbb{R} \,|\, t > U_m\}$ for $1 \le i \le m$, $1 \le j \le q$. Therefore, likelihood function (4.2.4) becomes

$$L(\beta, G) = \prod_{i=1}^{m} e^{\beta \mathbb{W}} (p_i)^{n\hat{\omega}_i} \Big(\sum_{k=i+1}^{m+1} p_k \Big)^{n\omega_i(\beta)} \equiv L(\beta, \boldsymbol{p}), \tag{4.3.1}$$

where

$$G(t) = \sum_{i=1}^{m} p_i I\{U_i \le t\}, \qquad \boldsymbol{p} = (p_1, \ldots, p_{m+1}) \tag{4.3.2}$$

satisfy

$$\begin{cases} p_i = G(U_i) - G(U_i-) = P_G(T = U_i), \quad \text{for } 1 \le i \le m \\ p_{m+1} = \bar{G}(U_m) = P_G(T > U_m) \\ \sum_{i=1}^{m+1} p_i = 1. \end{cases} \tag{4.3.3}$$

The reason we include p_{m+1} in (4.3.1)-(4.3.3) for d.f. $G(t)$ is that we may have the case $n\omega_m(\beta) > 0$ in (4.2.4), in which case any solution that maximizes $L(\beta, G)$ in (4.2.4) satisfies $\bar{G}(U_m) \equiv p_{m+1} > 0$. To see this, note that since $\beta > 0$ and $W_j > 0$ for all $1 \le j \le q$, we have

$$\omega_m(\beta) = \sum_{j=1}^{q} \hat{\omega}_{mj}(e^{\beta W_j} - 1) > \sum_{j=1}^{q} \hat{\omega}_{mj}(e^{\beta W_1} - 1) = (e^{\beta W_1} - 1)\sum_{j=1}^{q} \hat{\omega}_{mj} > 0,$$

whenever $\sum_{j=1}^{q} \hat{\omega}_{mj} = P_{\hat{F}_n}(T = U_m) > 0$.

The weighted empirical likelihood-based estimator $(\hat{\beta},\, \hat{G})$ for (β_0, G_0) is the solution that maximizes $L(\beta, G) = L(\beta, \boldsymbol{p})$ in (4.3.1) over all functions $G(t)$ in (4.3.2) satisfying (4.3.3). Specifically, $(\hat{\beta},\, \hat{G})$ is given by the solution of

$$\begin{cases} \max\ L(\beta, \boldsymbol{p}) = \prod_{i=1}^{m} e^{\beta \mathbb{W}} (p_i)^{n\hat{\omega}_i} \Big(\sum_{k=i+1}^{m+1} p_k \Big)^{n\omega_i(\beta)} \\ \text{subject to}: \quad 0 \le p_i \le 1, \quad 1 \le i \le m+1; \quad \sum_{i=1}^{m+1} p_i = 1. \end{cases} \tag{4.3.4}$$

The idea for solving (4.3.4) is outlined as follows:

Step 1: For fixed $\beta > 0$, find the solution $\hat{G}(\cdot\,;\, \beta)$ that maximizes $L(\beta, \boldsymbol{p})$;

Step 2: Obtain the profile likelihood function for β given by $\ell(\beta) = L(\beta, \hat{G}(\cdot\,;\, \beta))$, and find $\hat{\beta}$ that maximizes $\ell(\beta)$ for all $\beta > 0$. Then, the solution to optimization problem (4.3.4) is given by $(\hat{\beta}, \hat{G})$, where $\hat{G}(\cdot) = \hat{G}(\cdot\,; \hat{\beta})$.

Next, we follow Steps 1 and 2 to find the solution $(\hat{\beta}, \hat{G})$ to optimization problem (4.3.4).

Step 1: For fixed $\beta > 0$, since $e^{\beta \mathbb{W}}$ is independent of $\boldsymbol{p}$, we solve the following optimization problem:

$$\begin{cases} \max\ L_1(\boldsymbol{p}; \beta) \equiv \prod_{i=1}^{m} (p_i)^{n\hat{\omega}_i} \Big(\sum_{k=i+1}^{m+1} p_k \Big)^{n\omega_i(\beta)} \\ \text{subject to}: \quad 0 \le p_i \le 1, \quad 1 \le i \le m+1; \quad \sum_{i=1}^{m+1} p_i = 1. \end{cases} \tag{4.3.5}$$

In Section 4.5, we show that the solution to (4.3.5) is uniquely given by

$$\begin{cases} \hat{p}_i(\beta) = \hat{a}_i(\beta) \prod_{j=1}^{i-1} \big(1 - \hat{a}_j(\beta)\big), \quad 1 \le i \le m \\ \hat{a}_i(\beta) = \dfrac{n\hat{\omega}_i}{N_i(\beta)}, \qquad 1 \le i \le m, \end{cases} \tag{4.3.6}$$

with $\prod_{k=1}^{0} c_k \equiv 1$, and

$$L_1(\hat{\boldsymbol{p}}(\beta); \beta) = \prod_{i=1}^{m} \big(\hat{a}_i(\beta)\big)^{n\hat{\omega}_i} \big(1 - \hat{a}_i\beta)\big)^{N_i(\beta) - n\hat{\omega}_i}, \tag{4.3.7}$$

where for each $1 \le i \le m$, we define

$$n_i(\beta) \equiv n[\hat{\omega}_i + \omega_i(\beta)] \overset{(4.2.5)}{=} n \sum_{j=1}^{q} \hat{\omega}_{ij} e^{\beta W_j} \tag{4.3.8}$$

$$N_i(\beta) \equiv n_i(\beta) + \cdots + n_m(\beta) \ = \ \sum_{\ell=i}^{m} n_\ell(\beta) = n \sum_{\ell=i}^{m} \sum_{j=1}^{q} \hat{\omega}_{\ell j} e^{\beta W_j}. \tag{4.3.9}$$

Therefore, for any $U_1 \le t \le U_m$, we have

$$U_\alpha \le t \le U_{\alpha+1}, \quad \text{for some} \ \ 1 \le \alpha \le m-1, \tag{4.3.10}$$

which implies from (4.3.2) and (4.3.6),

$$
\begin{aligned}
\hat{G}(t;\beta) &= \sum_{i=1}^{m} \hat{p}_i(\beta) I\{U_i \le t\} = \sum_{i=1}^{m} \hat{p}_i(\beta) I\{U_i \le U_\alpha\} = \sum_{i=1}^{\alpha} \hat{p}_i(\beta) \\
&= \sum_{i=1}^{\alpha} \hat{a}_i(\beta) \prod_{j=1}^{i-1} \big(1-\hat{a}_j(\beta)\big) = \sum_{i=1}^{\alpha} \Big(1 - \big(1-\hat{a}_i(\beta)\big)\Big) \prod_{j=1}^{i-1} \big(1-\hat{a}_j(\beta)\big) \\
&= \sum_{i=1}^{\alpha} \Big(\prod_{j=1}^{i-1} \big(1-\hat{a}_j(\beta)\big) - \prod_{j=1}^{i} \big(1-\hat{a}_j(\beta)\big) \Big) \\
&= \Big(1 - \big(1-\hat{a}_1(\beta)\big)\Big) + \Big(\big(1-\hat{a}_1(\beta)\big) - \prod_{j=1}^{2} \big(1-\hat{a}_j(\beta)\big)\Big) + \cdots \\
&\quad + \Big(\prod_{j=1}^{\alpha-2} \big(1-\hat{a}_j(\beta)\big) - \prod_{j=1}^{\alpha-1} \big(1-\hat{a}_j(\beta)\big) \Big) \\
&\quad + \Big(\prod_{j=1}^{\alpha-1} \big(1-\hat{a}_j(\beta)\big) - \prod_{j=1}^{\alpha} \big(1-\hat{a}_j(\beta)\big) \Big) \\
&= 1 - \prod_{j=1}^{\alpha} \big(1-\hat{a}_j(\beta)\big) = 1 - \prod_{U_i \le U_\alpha} \big(1-\hat{a}_i(\beta)\big) = 1 - \prod_{U_i \le t} \big(1-\hat{a}_i(\beta)\big).
\end{aligned}
$$

Therefore, we have

$$
1 - \hat{G}(t;\beta) = \prod_{U_i \le t} \big(1-\hat{a}_i(\beta)\big), \tag{4.3.11}
$$

where $\hat{a}_i(\beta)$ is given by (4.3.6).

Step 2: From (4.3.4)-(4.3.7), the profile likelihood function is given by

$$
\begin{aligned}
\ell(\beta) &= e^{\beta \mathbb{W}} L_1(\hat{\boldsymbol{p}}(\beta);\beta) = e^{\beta \mathbb{W}} \prod_{i=1}^{m} \big(\hat{a}_i(\beta)\big)^{n\hat{\omega}_i} \big(1-\hat{a}_i(\beta)\big)^{N_i(\beta)-n\hat{\omega}_i} \\
&= e^{\beta \mathbb{W}} \prod_{i=1}^{m} \Big(\frac{n\hat{\omega}_i}{N_i(\beta)}\Big)^{n\hat{\omega}_i} \Big(1 - \frac{n\hat{\omega}_i}{N_i(\beta)}\Big)^{N_i(\beta)-n\hat{\omega}_i},
\end{aligned} \tag{4.3.12}
$$

which implies

$$\begin{aligned}
\log \ell(\beta) &= \beta\mathbb{W} + \sum_{i=1}^{m}\Bigg[n\hat{\omega}_i \log\Big(\frac{n\hat{\omega}_i}{N_i(\beta)}\Big) + \Big(N_i(\beta) - n\hat{\omega}_i\Big)\log\Big(\frac{N_i(\beta)-n\hat{\omega}_i}{N_i(\beta)}\Big)\Bigg] \\
&= \beta\mathbb{W} + \sum_{i=1}^{m}\Bigg[n\hat{\omega}_i\Big(\log n\hat{\omega}_i - \log N_i(\beta)\Big) \\
&\qquad + \Big(N_i(\beta) - n\hat{\omega}_i\Big)\Big(\log\big(N_i(\beta) - n\hat{\omega}_i\big) - \log N_i(\beta)\Big)\Bigg] \\
&= \beta\mathbb{W} + \sum_{i=1}^{m}\Bigg[n\hat{\omega}_i \log n\hat{\omega}_i + \big(N_i(\beta) - n\hat{\omega}_i\big)\log\big(N_i(\beta) - n\hat{\omega}_i\big) - N_i(\beta)\log N_i(\beta)\Bigg]. \quad (4.3.13)
\end{aligned}$$

Taking the derivative of (4.3.13), we have

$$\begin{aligned}
\frac{d}{d\beta}\Big(\log \ell(\beta)\Big) &= \mathbb{W} + \sum_{i=1}^{m}\Bigg[\big(N_i(\beta) - n\hat{\omega}_i\big)\frac{N_i'(\beta)}{N_i(\beta) - n\hat{\omega}_i} + N_i'(\beta)\log\big(N_i(\beta) - n\hat{\omega}_i\big) \\
&\qquad - N_i(\beta)\frac{N_i'(\beta)}{N_i(\beta)} - N_i'(\beta)\log N_i(\beta)\Bigg] \\
&= \mathbb{W} + \sum_{i=1}^{m}\Bigg[N_i'(\beta)\log\big(N_i(\beta) - n\hat{\omega}_i\big) - N_i'(\beta)\log N_i(\beta)\Bigg] \\
&= \mathbb{W} + \sum_{i=1}^{m} N_i'(\beta)\log\frac{N_i(\beta) - n\hat{\omega}_i}{N_i(\beta)} = \mathbb{W} + \sum_{i=1}^{m} N_i'(\beta)\log\Bigg(1 - \frac{n\hat{\omega}_i}{N_i(\beta)}\Bigg) \\
&\overset{(4.2.5)}{=} \sum_{j=1}^{q} W_j + \sum_{i=1}^{m} N_i'(\beta)\log\Bigg(1 - \frac{n\hat{\omega}_i}{N_i(\beta)}\Bigg), \quad (4.3.14)
\end{aligned}$$

where from (4.3.9), we know

$$N_i'(\beta) = \frac{d}{d\beta}\Bigg(n\sum_{\ell=i}^{m}\sum_{j=1}^{q}\hat{\omega}_{\ell j}e^{\beta W_j}\Bigg) = n\sum_{\ell=i}^{m}\sum_{j=1}^{q}W_j\hat{\omega}_{\ell j}e^{\beta W_j}. \quad (4.3.15)$$

Thus, estimator $\hat{\beta}$ for β_0 for Cox model (1.2.24) based on weighted empirical likelihood function (4.2.4) with censored data (4.1.1) is given by a solution of the following estimating equation:

$$\sum_{j=1}^{q} W_j + n\sum_{i=1}^{m}\Bigg(\sum_{\ell=i}^{m}\sum_{j=1}^{q}W_j\hat{\omega}_{\ell j}e^{\beta W_j}\Bigg)\log\Bigg(1 - \frac{\sum_{j=1}^{q}\hat{\omega}_{ij}}{\sum_{\ell=i}^{m}\sum_{j=1}^{q}\hat{\omega}_{\ell j}e^{\beta W_j}}\Bigg) = 0. \quad (4.3.16)$$

Therefore, estimator $(\hat{\beta}, \hat{G})$ for (β_0, G_0) in Cox model (1.2.24) with censored data (4.1.1) is given by $\hat{G}(t) = \hat{G}(t; \hat{\beta})$ as in (4.3.11) with $\hat{\beta}$ as the solution of estimating equation (4.3.16).

To compute $\hat{\beta}$, the Newton-Raphson method may be used with Cox's partial likelihood estimate $\hat{\beta}_c$ as the initial value. This is to be further studied in the future.

4.4 Remarks

Remark 4.1: As reviewed in Section 1.3.5, currently there has not been any published work on the Cox model with doubly censored data (1.3.6) or with partly interval-censored data (1.3.12)-(1.3.13). Here, our work provides solutions to these problems in a unified form, because our results in Sections 4.2 and 4.3 hold for any type of censored data whose estimator $\hat{F}_n(t, z)$ for d.f. $F_0(t, z)$ can be expressed as (4.1.2). It should be noted that equation (4.3.16) is relatively easy to solve with the use of a computer, where only one program needs to be written to find the weighted empirical likelihood-based estimators $(\hat{\beta}, \hat{G})$ for (β_0, G_0) for Cox model (1.2.24) with various types of censored data.

Remark 4.2: In this work we considered Cox model (1.2.24) in which there is only one explanatory variable Z. It should be noted that the extension of these results to multivariate explanatory variables $\boldsymbol{Z} = (\mathcal{Z}_1, \ldots, \mathcal{Z}_k)$ in Cox model (1.2.22) is straightforward, and it works for various types of censored data mentioned in Section 1.3.

4.5 Proofs

Proof of (4.3.6)-(4.3.7): To find the solution to (4.3.5), we consider the following substitutions:

$$a_i = \frac{p_i}{b_i} \qquad \text{and} \qquad b_i = \sum_{k=i}^{m+1} p_k, \qquad 1 \leq i \leq m, \tag{4.5.1}$$

which imply

$$b_{m+1} = p_{m+1}, \qquad b_{i+1} = \sum_{k=i+1}^{m+1} p_k \; = \; b_i - p_i, \qquad \text{and} \qquad 1 - a_i = \frac{b_{i+1}}{b_i}. \tag{4.5.2}$$

Here, we follow the same procedure used in the proof of Theorem 2.1(iii). From (4.5.1)-(4.5.2) and (4.3.8)-(4.3.9), likelihood function (4.3.5) can be equivalently written as

$$\begin{aligned}
L_1(\boldsymbol{p};\beta) &= \prod_{i=1}^{m}(p_i)^{n\hat{\omega}_i}(b_{i+1})^{n\omega_i(\beta)} \; = \; \prod_{i=1}^{m}(p_i)^{n\hat{\omega}_i}(b_i - p_i)^{n\omega_i(\beta)} \\
&= \prod_{i=1}^{m}(p_i)^{n\hat{\omega}_i}\Big(\frac{p_i}{a_i} - p_i\Big)^{n\omega_i(\beta)} \; = \; \prod_{i=1}^{m}(p_i)^{n[\hat{\omega}_i+\omega_i(\beta)]}\Big(\frac{1-a_i}{a_i}\Big)^{n\omega_i(\beta)} \\
&= \Big(\prod_{i=1}^{m}(p_i)^{n_i(\beta)}\Big)\prod_{i=1}^{m}\frac{(1-a_i)^{n_i(\beta)-n\hat{\omega}_i}}{(a_i)^{n_i(\beta)-n\hat{\omega}_i}} \\
&= \Big(\prod_{i=1}^{m}(p_i)^{n_i(\beta)}\Big)\frac{\prod_{i=1}^{m}(a_i)^{n\hat{\omega}_i}(1-a_i)^{n-n\hat{\omega}_i-(n_1(\beta)+\cdots+n_{i-1}(\beta))}}{\prod_{i=1}^{m}(a_i)^{n_i(\beta)}(1-a_i)^{n-(n_1(\beta)+\cdots+n_i(\beta))}} \\
&= \Big(\prod_{i=1}^{m}(p_i)^{n_i(\beta)}\Big)\frac{\prod_{i=1}^{m}(a_i)^{n\hat{\omega}_i}(1-a_i)^{n-n\hat{\omega}_i-(N_1(\beta)-N_i(\beta))}}{\prod_{i=1}^{m}(a_i)^{n_i(\beta)}(1-a_i)^{n-(N_1(\beta)-N_{i+1}(\beta))}} \\
&= \Big(\prod_{i=1}^{m}(p_i)^{n_i(\beta)}\Big)\frac{\prod_{i=1}^{m}(a_i)^{n\hat{\omega}_i}(1-a_i)^{n-n\hat{\omega}_i-(N_1(\beta)-N_i(\beta))}}{\prod_{i=1}^{m}\Big(\frac{p_i}{b_i}\Big)^{n_i(\beta)}(1-a_i)^{n-(N_1(\beta)-N_{i+1}(\beta))}} \\
&= \Big(\prod_{i=1}^{m}(b_i)^{n_i(\beta)}\Big)\frac{\prod_{i=1}^{m}(a_i)^{n\hat{\omega}_i}(1-a_i)^{n-n\hat{\omega}_i-(N_1(\beta)-N_i(\beta))}}{\prod_{i=1}^{m}(1-a_i)^{n-(N_1(\beta)-N_{i+1}(\beta))}}.
\end{aligned} \tag{4.5.3}$$

From (4.5.2), (4.3.12) and $b_1 = \sum_{i=1}^{m+1} p_i = 1$, the denominator in (4.5.3) is equivalent to

$$\begin{aligned}
\prod_{i=1}^{m}(1-a_i)^{n-(N_1(\beta)-N_{i+1}(\beta))} \; &= \; \prod_{i=1}^{m}\Big(\frac{b_{i+1}}{b_i}\Big)^{n-(N_1(\beta)-N_{i+1}(\beta))} \\
&= \frac{(b_2)^{n-n_1(\beta)}}{(b_1)^{n-n_1(\beta)}}\;\frac{(b_3)^{n-n_1(\beta)-n_2(\beta)}}{(b_2)^{n-n_1(\beta)-n_2(\beta)}}\cdots\cdot\frac{(b_{m+1})^{n-n_1(\beta)-\cdots-n_m(\beta)}}{(b_m)^{n-n_1(\beta)-\cdots-n_m(\beta)}} \\
&= \frac{1}{(b_1)^n}\Big((b_1)^{n_1(\beta)}(b_2)^{n_2(\beta)}\cdots\cdot(b_m)^{n_m(\beta)}\Big)(b_{m+1})^{n-N_1(\beta)} \\
&= \Big(\prod_{i=1}^{m}(b_i)^{n_i(\beta)}\Big)\frac{(b_{m+1})^{n-N_1(\beta)}}{(b_1)^n} = \Big(\prod_{i=1}^{m}(b_i)^{n_i(\beta)}\Big)(b_{m+1})^{n-N_1(\beta)},
\end{aligned} \tag{4.5.4}$$

where the last equality is true because $b_1 = \sum_{i=1}^{m+1} p_i = 1$. Since $n - N_1(\beta)$ is independent of i, we have from (4.5.2)

$$\begin{aligned} \prod_{i=1}^{m}(1-a_i)^{n-N_1(\beta)} &= \Big(\prod_{i=1}^{m}(1-a_i)\Big)^{n-N_1(\beta)} \quad = \quad \Big(\prod_{i=1}^{m}\frac{b_{i+1}}{b_i}\Big)^{n-N_1(\beta)} \\ &= \Big(\frac{b_2}{b_1}\,\frac{b_3}{b_2}\cdots\cdot\frac{b_{m+1}}{b_m}\Big)^{n-N_1(\beta)} \quad = \quad (b_{m+1})^{n-N_1(\beta)} \end{aligned} \tag{4.5.5}$$

and plugging this into (4.5.4), we obtain

$$\prod_{i=1}^{m}(1-a_i)^{n-(N_1(\beta)-N_{i+1}(\beta))} \quad = \quad \Big(\prod_{i=1}^{m}(b_i)^{n_i(\beta)}\Big)\prod_{i=1}^{m}(1-a_i)^{n-N_1(\beta)}. \tag{4.5.6}$$

Plugging (4.5.6) into (4.5.3), the likelihood function becomes

$$\begin{aligned} L_1(\boldsymbol{p};\beta) &= \frac{\prod_{i=1}^{m}(a_i)^{n\hat{\omega}_i}(1-a_i)^{n-n\hat{\omega}_i-(N_1(\beta)-N_i(\beta))}}{\prod_{i=1}^{m}(1-a_i)^{n-N_1(\beta)}} \\ &= \prod_{i=1}^{m}(a_i)^{n\hat{\omega}_i}(1-a_i)^{N_i(\beta)-n\hat{\omega}_i} \equiv \mathcal{G}(\boldsymbol{a};\beta), \end{aligned} \tag{4.5.7}$$

where $\boldsymbol{a} = (a_1,\ldots,a_m)$ with a_i's given by (4.5.1). Therefore, optimization problem (4.3.5) is equivalent to the following optimization problem:

$$\begin{cases} \max\, \mathcal{G}(\boldsymbol{a};\beta) = \prod_{i=1}^{m}(a_i)^{n\hat{\omega}_i}(1-a_i)^{N_i(\beta)-n\hat{\omega}_i} \\ \text{subject to :} \quad 0 \le a_i \le 1, \quad \text{for } 1 \le i \le m. \end{cases} \tag{4.5.8}$$

Note that the set

$$\big\{\boldsymbol{a}\,|\,0 \le a_i \le 1,\ 1 \le i \le m\big\}$$

is compact and $\mathcal{G}(\boldsymbol{a};\beta)$ is continuous in $\boldsymbol{a}$, thus a solution to (4.5.8) exists from Weierstrass' Theorem (Bazaraa *et al.*, 1993; page 41). Let $\hat{\boldsymbol{a}}$ denote the solution to optimization problem (4.5.8). Suppose $n\hat{\omega}_i > 0$ and $N_i(\beta) - n\hat{\omega}_i > 0$. Therefore, any solution $\hat{\boldsymbol{a}}$ that maximizes $\mathcal{G}(\boldsymbol{a};\beta)$ in (4.5.8) will satisfy $0<\hat{a}_i<1$. Therefore, optimization problem (4.5.8) is equivalent to

$$\begin{cases} \max\, \log\mathcal{G}(\boldsymbol{a};\beta) = \displaystyle\sum_{i=1}^{m} \Big[n\hat{\omega}_i \log a_i + (N_i(\beta) - n\hat{\omega}_i)\log(1-a_i) \Big] \\ \text{subject to :} \quad 0<a_i<1, \quad 1\le i\le m. \end{cases} \tag{4.5.9}$$

Note that $\log\mathcal{G}(\boldsymbol{a};\beta)$ is well-defined on the open convex set

$$A \equiv \big\{ \boldsymbol{a} \,|\, 0<a_i<1,\ 1\le i\le m \big\},$$

and from $\partial \log\mathcal{G}(\boldsymbol{a};\beta)/\partial a_i = 0$, we have for $1\le i\le m$

$$\frac{\partial}{\partial a_i}\Big(\log\mathcal{G}(\boldsymbol{a};\beta)\Big) = \frac{n\hat{\omega}_i}{a_i} - \frac{N_i(\beta)-n\hat{\omega}_i}{1-a_i} = 0 \qquad \Longleftrightarrow \qquad \hat{a}_i(\beta) = \frac{n\hat{\omega}_i}{N_i(\beta)}.$$

Note that $\hat{\boldsymbol{a}}(\beta) = \{\hat{a}_i(\beta)\,|\,1\le i\le m\}$ is a feasible solution for optimization problem (4.5.9) because $n\hat{\omega}_i>0$ and $N_i(\beta)-n\hat{\omega}_i>0$ imply $0<\hat{a}_i<1$. Since for all $\boldsymbol{a}\in A$, we have

$$\frac{\partial^2}{\partial a_i^2}\Big(\log\mathcal{G}(\boldsymbol{a};\beta)\Big) \;=\; \frac{-n\hat{\omega}_i}{(a_i)^2} - \frac{N_i(\beta)-n\hat{\omega}_i}{(1-a_i)^2} \;<\; 0, \tag{4.5.10}$$

$$\frac{\partial^2}{\partial a_i\partial a_k}\Big(\log\mathcal{G}(\boldsymbol{a};\beta)\Big) \;=\; 0 \qquad\qquad \text{for } i\neq k, \tag{4.5.11}$$

we see that the Hessian matrix for $\log\mathcal{G}(\boldsymbol{a};\beta)$ will have diagonal entries given by (4.5.10) and 0's elsewhere, which implies $\log\mathcal{G}(\boldsymbol{a};\beta)$ is strictly concave down on set A. Therefore, the solution for (4.5.9) is uniquely given by

$$\hat{a}_i(\beta) = \frac{n\hat{\omega}_i}{N_i(\beta)}, \qquad 1\le i\le m. \tag{4.5.12}$$

To get $\hat{p}_i(\beta)$ in terms of $\hat{a}_i(\beta)$, note that from (4.5.1)

$$
\begin{aligned}
\hat{p}_1(\beta) &= \hat{a}_1(\beta)\hat{b}_1(\beta) = \hat{a}_1(\beta)\sum_{j=1}^{m+1}\hat{p}_j(\beta) = \hat{a}_1(\beta) \\
\hat{p}_2(\beta) &= \hat{a}_2(\beta)\hat{b}_2(\beta) = \hat{a}_2(\beta)\sum_{j=2}^{m+1}\hat{p}_j(\beta) = \hat{a}_2(\beta)\big(1-\hat{p}_1(\beta)\big) \\
&= \hat{a}_2(\beta)\big(1-\hat{a}_1(\beta)\big), \\
\hat{p}_3(\beta) &= \hat{a}_3(\beta)\hat{b}_3(\beta) = \hat{a}_3(\beta)\sum_{j=3}^{m+1}\hat{p}_j(\beta) = \hat{a}_3(\beta)\big(1-\hat{p}_1(\beta)-\hat{p}_2(\beta)\big) \\
&= \hat{a}_3(\beta)\big(1-\hat{a}_1(\beta)-\hat{a}_2(\beta)+\hat{a}_1(\beta)\hat{a}_2(\beta)\big) \\
&= \hat{a}_3(\beta)\big(1-\hat{a}_2(\beta)\big)-\hat{a}_1(\beta)\hat{a}_3(\beta)\big(1-\hat{a}_2(\beta)\big) \\
&= \hat{a}_3(\beta)\big(1-\hat{a}_1(\beta)\big)\big(1-\hat{a}_2(\beta)\big) = \hat{a}_3(\beta)\prod_{j=1}^{2}(1-\hat{a}_j(\beta)).
\end{aligned}
$$

Continue this and we get a general expression for $\hat{p}_i(\beta)$ as follows:

$$
\hat{p}_i(\beta) = \hat{a}_i(\beta)\prod_{j=1}^{i-1}\big(1-\hat{a}_j(\beta)\big), \quad 1 \le i \le m. \tag{4.5.13}
$$

where $\prod_{k=1}^{0} c_k$ is set as 1 when it occurs. $\square$

CHAPTER 5. CONCLUDING REMARKS

In Chapter 2, we provided the bivariate nonparametric maximum likelihood estimator (BNPMLE) $\hat{F}_n(t,z)$ for the bivariate distribution function $F_0(t,z)$ based on right censored survival data (2.1.2) in which the survival time T is subject to right censoring and the covariate Z is a scalar and is completely observable. This BNPMLE $\hat{F}_n$ provides a completely nonparametric method for the data analysis on the studies of the relation between T and Z. But the asymptotic properties of $\hat{F}_n$ needs to be studied in the future.

In Chapter 3, we derived empirical likelihood based confidence interval for conditional survival probabilities with right censored data (2.1.2), and we provided an analytic expression for the empirical likelihood ratio. The asymptotic distribution of such likelihood ratio will be studied in the future.

In Chapter 4, we derived the estimator $(\hat{\beta}, \hat{G})$ for (β_0, G_0) in Cox model (1.2.24) with general censored data (4.1.1). Our methods here hold for various types of censored data, including some of those that have not been previously studied in the literature, such as doubly censored data (1.3.6) and party interval-censored data (1.3.12)-(1.3.13). But the asymptotic properties of estimator $(\hat{\beta}, \hat{G})$ needs to be studied in the future.

LIST OF REFERENCES

Andersen, P.K. and Gill, R.D. (1982). Cox's Regression Model for Counting Processes: A Large Sample Study. *Annals of Statistics* **10**, 1100-1120.

Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, 1st Edition. Springer.

Bazaraa, M.S., Sherali, H.D. and Shetty, C.M. (1993). *Nonlinear Programming: Theory and Algorithms*, 2nd Edition. John Wiley & Sons, INC.

Casella, G. and Berger, R.L. (2002). *Statistical Inference*, 2nd Edition. Duxbury.

Chang, M.N. and Yang, G.L. (1987). Strong Consistency of a Nonparametric Estimator of the Survival Function with Doubly Censored Data. *Annals of Statistics* **15**, 1536-1547.

Cox, D.R. (1972). Regression Models and Life Tables (with Discussion). *Journal of the Royal Statistical Society, Series B* **34**, 187-221.

Cox, D.R. and Oakes, D. (1984). *Analysis of Survival Data.* Chapman & Hall.

Cox, D.R. and Snell, E.J. (1968). A General Definition of Residuals. *Journal of the Royal Statistical Society, Series B* **30**, 248-275.

Cupples, L.A., Risch, N., Farrer, L.A. and Myers, R.H. (1991). Estimation of Morbid Risk and Age at Onset with Missing Information. *American Journal of Human Genetics* **49**, 76-87.

Dabrowska, D.M. (1988). Kaplan-Meier Estimate on the Plane. *Annals of Statistics* **16**, 1475-1489.

Dabrowska, D.M. (1989). Uniform Consistency of the Kernel Conditional Kaplan-Meier Estimate. *Annals of Statistics* **17**, 1157-1167.

De Gruttola, V. and Lagakos, S.W. (1989). Analysis of Doubly Censored Survival Data, with Application to AIDS. *Biometrics* **45**, 1-11.

DiCiccio, T.J., Hall, P.J. and Romano, J. (1991). Empirical Likelihood is Bartlett-correctable. *Annals of Statistics* **19**, 1053-1061.

Efron, B. (1967). The Power of the Likelihood Ratio Test. *Annals of Mathematical Statistics* **38**, 802-806.

Efron, B. (1977). The Efficiency of Cox's Likelihood Function for Censored Data. *Journal of the American Statistical Association* **72**, 557-565.

Efron, B. and Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science* **1**, 54-75.

Farrington, C.P. (2000). Residuals for Proportional Hazards Models with Interval Censored Survival Data. *Biometrics* **56**, 473-482.

Feinleib, M., Kannel, W.B., Garrison, R.J., McNamara, P.M. and Castelli, W.P. (1975). The Framingham Offspring Study: Design and Preliminary Data. *Preventive Medicine* **4**, 518-525.

Gehan, E.A. (1965). A Generalized Wilcoxen Test for Comparing Arbitrarily Single-Censored Samples. *Biometrika* **52**, 203-224.

Gill, R.D. (1983). Large Sample Behavior of the Product-Limit Estimator on the Whole Line. *Annals of Statistics* **11**, 49-58.

Groeneboom, P. and Wellner, J.A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* Birkhäuser-Verlag.

Gu, M.G. and Zhang, C.H. (1993). Asymptotic Properties of Self-Consistent Estimators Based On Doubly Censored Data. *Annals of Statistics* **21**, 611-624.

Hald, A. (1949). Maximum Likelihood Estimation of the Parameters of a Normal Distribution Which is Truncated at a Known Point. *Skand. Atkuarietidskr* **32**, 119-134.

Hogg, R.V. and Craig, A.T. (1995). *Introduction to Mathematical Statistics*, 5th Edition. Prentice-Hall, INC.

Huang, J. (1996). Efficient Estimation for the Proportional Hazards Model with Interval Censoring. *Annals of Statistics* **24**, 540-568.

Huang, J. (1999). Asymptotic Properties of Nonparametric Estimation Based on Partly Interval-Censored Data. *Statistica Sinica* **9**, 501-519.

Huang, J. and Wellner, J.A. (1995). Efficient Estimation for the Cox Model With Case 2 Interval Censoring. Technical Report, Department of Statistics, University of Washington.

Huang, J. and Wellner, J.A. (1997). *Interval Censored Survival Data: A Review of Recent Progress.* Springer-Verlag, INC.

Kalbfleisch, J.D. and Prentice, R.L. (1973). Marginal Likelihoods Based on Cox's Regression and Life Model, *Biometrika*, **60**, 267-278.

Kalbfleisch, J.D. and Prentice, R.L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd Edition. John Wiley & Sons, INC.

Kaplan, E.L. and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association* **53**, 457-481.

Kardaun, O. (1983). Statistical Analysis of Male Larynx-Cancer Patients-A Case Study. *Statistica Neerlandica* **37**, 103-126.

Kim, M.Y., De Gruttola, V. and Lagakos, S.W. (1993). Analyzing Doubly Censored Data with Covariates, with Applications to AIDS. *Biometrics* **49**, 13-22.

Klein, J.P. and Moeschberger, M.L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data.* Springer-Verlag, INC.

Lagakos, S.W. (1979). General Right Censoring and Its Impact on the Analysis of Survival Data. *Biometrics* **35**, 139-156.

Lagakos, S.W. (1980). The Graphical Evaluation of Explanatory Variables in Proportional Hazard Regression Models. *Biometrika* **68**, 93-98.

Lin, D.Y. and Ying, Z. (1993). A Simple Nonparametric Estimator of the Bivariate Survival Function Under Univariate Censoring. *Biometrika* **80**, 573-581.

Mykland, P.A. (1995). Dual Likelihood. *Annals of Statistics* **23**, 396-421.

Mykland, P.A. and Ren, J. (1996). Algorithms for Computing Self-Consistent and Maximum Likelihood Estimators with Doubly Censored Data. *Annals of Statistics* **24**, 1740-1764.

Oakes, D. (2000). Survival Analysis. *Journal of the American Statistical Association* **95**, 282-285.

Oakes, D. (2001). Biometrika Centenary: Survival Analysis. *Biometrika* **88**, 99-142.

O'Brien, T.R., Busch, M.P., Donegan, E., Ward, J.W., Wong, L., Samson, S.M., Perkins, H.A., Altman, R., Stoneburner, R.L. and Holmberg, S.D. (1994). Heterosexual Transmission of Human Immunodeficiency Virus Type 1 from Transfusion Recipients to their Sex Partners. *Journal of Acquired Immune Deficiency Syndromes* **7**, 705-710.

Odell, P.M., Anderson, K.M. and D'Agostino, R.B. (1992). Maximum Likelihood Estimation for Interval Censored Data Using a Weibull-Based Accelerated Failure Time Model. *Biometrics* **48**, 951-959.

Owen, A.B. (1988). Empirical Likelihood Ratio Confidence Intervals For a Single Functional. *Biometrika* **75**, 237-249.

Owen, A.B. (1990). Empirical Likelihood Ratio Confidence Regions. *Annals of Statistics* **18**, 90-120.

Owen, A.B. (1991). Empirical Likelihood for Linear Models. *Annals of Statistics* **19**, 1725-1747.

Owen, A.B. (2001). *Empirical Likelihood.* Chapman & Hall/CRC.

Pan, W. (1999). Extending the Iterative Convex Minorant Algorithm to the Cox Model for Interval Censored Data. *Journal of Computational and Graphical Statistics* **8**, 109-120.

Pan, W. (2001). A Multiple Imputation Approach to Regression Analysis for Doubly Censored Data with Application to AIDS Studies. *Biometrics* **57**, 1245-1250.

Peer, P.G., Van Dijck, J.A., Hendriks, J.H., Holland, R. and Verbeek A.L. (1993). Age-Dependent Growth Rate of Primary Breast Cancer. *Cancer* **71**, 3547-3551.

Qin, J. and Lawless, J. (1994). Empirical Likelihood and General Estimating Equations. *Annals of Statistics* **22**, 300-325.

Ren, J. (2001). Weighted Empirical Likelihood Ratio Confidence Intervals for the Mean with Censored Data. *Annals of the Institute of Statistical Mathematics* **53**, 498-516.

Ren, J. (2003). Goodness of Fit Tests with Interval Censored Data. *Scandinavian Journal of Statistics* **30**, 211-226.

Ren, J. (2008a). Weighted Empirical Likelihood in Some Two-Sample Semiparametric Models with Various Types of Censored Data. *Annals of Statistics* **36**, 147-166)

Ren, J. (2008b). Smoothed Weighted Empirical Likelihood Ratio Confidence Intervals for Quantiles. *Bernoulli* **14**, 725-748.

Ren, J. and Gu, M.G. (1997). Regression M-Estimators with Doubly Censored Data. *Annals of Statistics* **25**, 2638-2664.

Ren, J. and He, B. (2011). Estimation and Goodness-of-Fit for the Cox Model with Various Types of Censored Data. Journal of Statistical Planning Inference **141**, 961-971.

Ren, J. and Riddlesworth, T. (2011). Bivariate Nonparametric Maximum Likelihood Estimator with Right Censored Data. (in preparation).

Ren, J. and Peer, P.G. (2000). A Study on Effectiveness of Screening Mammograms. *International Journal of Epidemiology* **29**, 803-806.

Ren, J. and Zhou, M. (2011). Full Likelihood Inference in the Cox Model: An Empirical Likelihood Approach. *Annals of the Institute of Statistical Mathematics* **63**, 1005-1018.

Royden, H.L. (1988). *Real Analysis*, 3rd Edition. Prentice-Hall, INC.

Satten, G.A. (1996). Rank-Based Inference in the Proportional Hazards Model for Interval Censored Data. *Biometrika* **83**, 355-370.

Schoenfeld, D. (1982). Partial Residuals for the Proportional Hazards Regression Model. *Biometrika* **69**, 239-241

Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics.* John Wiley & Sons, INC.

Stute, W. and Wang, J.L. (1993). The Strong Law Under Random Censorship. *Annals of Statistics* **21**, 1591-1607.

Sun, J., Liao, Q. and Pagano, M. (1999). Regression Analysis of Doubly Censored Failure Time Data with Applications to AIDS Studies. *Biometrics* **55**, 909-914.

Therneau, T.M., Grambsch, P.M. and Fleming, T.R. (1990). Martingale-Based Residuals for Survival Models. *Biometrika* **77**, 147-160.

Tsiatis, A.A. (1981). A Large Sample Study of Cox's Regression Model. *Annals of Statistics* **9**, 93-108.

Turnbull, B.W. (1974). Nonparametric Estimation of a Survivorship Function with Doubly Censored Data. *Journal of the American Statistical Association* **69**, 169-173.

Wilks, S.S. (1938). The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *Annals of Mathematical Statistics* **9**, 60-62.