STARS

Electronic Theses and Dissertations, 2004-2019

2013

Accelerated Life Model With Various Types Of Censored Data

Kathryn Pridemore
University of Central Florida

University of Central Florida

STARS
Showcase of Text, Archives, Research & Scholarship

ACCELERATED LIFE MODEL WITH VARIOUS TYPES OF CENSORED DATA

by

KATIE PRIDEMORE
B.S. Mathematics, Stetson University, 2003
M.S. Mathematics, University of Central Florida, 2005

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

Summer Term
2013

Major Professor:
Marianna Pensky

ABSTRACT

The Accelerated Life Model is one of the most commonly used tools in the analysis of survival data which are frequently encountered in medical research and reliability studies. In these types of studies we often deal with complicated data sets for which we cannot observe the complete data set in practical situations due to censoring. Such difficulties are particularly apparent by the fact that there is little work in statistical literature on the Accelerated Life Model for complicated types of censored data sets, such as doubly censored data, interval censored data, and partly interval censored data.

In this work, we use the *Weighted Empirical Likelihood* approach (Ren, 2001) [33] to construct tests, confidence intervals, and goodness-of-fit tests for the Accelerated Life Model in a unified way for various types of censored data. We also provide algorithms for implementation and present relevant simulation results.

I began working on this problem with Dr. Jian-Jian Ren. Upon Dr. Ren's departure from the University of Central Florida I completed this dissertation under the supervision of Dr. Marianna Pensky.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

The *Accelerated Life Model* is one of the most commonly used tools in the analysis of survival data. Due to the nature of survival data, we often encounter data sets which are subject to censoring, i.e., we cannot observe the complete data set in practical situations. Until now, there has been little work in statistical literature on the Accelerated Life Model for complicated types of censored data sets, such as doubly censored data, interval censored data, and partly interval censored data. In this research, we use the *Weighted Empirical Likelihood* approach (Ren, 2001) [33] to construct tests, confidence intervals, and goodness-of-fit tests for the Accelerated Life Model in a unified way for various types of censored data.

This chapter is organized as follows. Section 1.1 briefly introduces some basic concepts and notations in survival analysis. Section 1.2 introduces the Accelerated Life Model and reviews some relevant recent works. Section 1.3 describes various types of censored data with examples, and reviews some relevant asymptotic results on the nonparametric maximum likelihood distribution estimators. Section 1.4 reviews the techniques of Parametric Likelihood, Empirical Likelihood (Owen, 1988)[31], and Weighted Empirical Likelihood (Ren, 2001)[33]. Finally, Section 1.6 summarizes the main results of this dissertation, and outlines the organization of the rest of this dissertation.

1.1 Introduction

Survival analysis is an area of statistical research which is concerned with the failure time of subjects. An example of this in medical research is a treatment study where a new drug or treatment is being tested. Researchers may want to determine the effects of the new treatment on survival time for patients with diseases such as diabetes, AIDS, cancer, etc.

While there may be many different explanatory variables, survival analysis primarily deals with a univariate lifetime variable, which is often referred to as *failure time*. To determine failure time precisely, we need a clearly defined time origin, a way of measuring time, and an explicit definition of the meaning of failure. In medical research, the time origin is usually defined as the time at which a patient enters a clinical trial, time is measured in days or months, and failure is defined as the time when a disease relapses or the time when a patient dies from the disease of interest.

One challenge in the analysis of survival data is that in practical situations, we often are unable to observe the failure time of an individual due to censoring. Such a challenge can be quite difficult to handle mathematically, which is why there has been so little work done in the statistical literature on the Accelerated Life Model with complicated types of censored data, such as doubly censored, interval censored, and partly interval censored data. But it is well known that these complicated types of censored data are encountered in important clinical trials in medical research; see Section 1.3 for descriptions of various types of censored data

and real data examples. Thus, it is important for us to develop new statistical procedures to handle these types of censored data.

As follows, we introduce some commonly used definitions and notations in survival analysis. Let T denote the lifetime random variable, which is continuous and nonnegative. And let $f_T(t)$ and $F_T(t)$ denote the density function and distribution function of T, respectively.

Definition 1.1. The *survival function* of T is defined by

$$\bar{F}_T(t) \; = \; P\{T \geq t\} \; = \; 1 - F_T(t). \tag{1.1}$$

Definition 1.2. The *hazard function* of T is defined by

$$h_T(t) = \lim_{\Delta \to 0+} \frac{P\{t \leq T < t + \Delta \mid T \geq t\}}{\Delta}. \tag{1.2}$$

Note that the hazard function is the instantaneous rate of mortality at time t given $T \geq t$. Also, note that Definition 1.2 implies that the hazard function can be expressed in terms of the density function and distribution function of T as below:

$$\begin{aligned}
h_T(t) &= \lim_{\Delta \to 0+} \frac{P\{t \leq T < t + \Delta \mid T \geq t\}}{\Delta} = \lim_{\Delta \to 0+} \frac{P\{t \leq T < t + \Delta\}}{P\{T \geq t\} \cdot \Delta} \\
&= \lim_{\Delta \to 0+} \frac{F_T(t + \Delta) - F_T(t)}{\bar{F}_T(t) \cdot \Delta} = \frac{F_T'(t)}{\bar{F}_T(t)} = \frac{f_T(t)}{\bar{F}_T(t)}.
\end{aligned} \tag{1.3}$$

The hazard function $h_T(t)$ in (1.2) plays a key role in the *Cox Proportional Hazards Model* (Cox, 1972) [8], which is one of the most commonly used models in survival analysis. However, the model assumptions of the Cox Proportional Hazards Model do not hold in some practical situations. Hence, the Accelerated Life Model is a commonly used alternative model in survival analysis. In the next section, we describe the Accelerated Life Model and discuss its relation to the Cox Proportional Hazards Model.

1.2 Accelerated Life Model

In this section, we describe the *Accelerated Life Model* (ALM). Specifically, Subsection 1.2.1 discusses the Two-Sample Accelerated Life Model; Subsection 1.2.2 discusses the general case of the Accelerated Life Model and its relationship with the Cox Proportional Hazards model; and Subsection 1.2.3 briefly reviews some recent relevant works on the Accelerated Life Model.

1.2.1 Two-Sample Accelerated Life Model

The basic underlying assumption for the Accelerated Life Model is that the covariate vector z acts multiplicatively on the failure times. The following summarizes the relevant topics from Section 5.1 of Cox and Oakes (1984). [9]

Consider the simple case where we have a single indicator variable z such that $z = 0$ corresponds to the control group, and $z = 1$ corresponds to the treatment group. In survival analysis, this could mean a study which assesses the effectiveness of a new medical treatment that is anticipated to increase the survival time of individuals in, say, a cancer study. The data from the treatment group $z = 1$ and the control group $z = 0$ are the following lifetime random samples, respectively:

$$\textbf{Treatment } \boldsymbol{z = 1}: \quad X_1, \ldots, X_{n_1} \overset{i.i.d.}{\sim} F_X,$$
$$\textbf{Control } \boldsymbol{z = 0}: \quad Y_1, \ldots, Y_{n_0} \overset{i.i.d.}{\sim} F_Y. \tag{1.4}$$

4

The assumption of the Accelerated Life Model is that lifetime random variables X and Y are proportional to each other, which is denoted by

$$Y = \frac{X}{\gamma_0}, \tag{1.5}$$

where γ_0 is an unknown positive *scale parameter*, and model (1.5) is referred to as the *Two-Sample Accelerated Life Model*.

From model assumption (1.5), the following relationships among the distribution function, density function, and hazard function of the random variables X and Y are implied:

$$F_Y(t) = P\{Y \le t\} = P\left\{\frac{X}{\gamma_0} \le t\right\} = P\{X \le \gamma_0 t\} = F_X(\gamma_0 t),$$

$$f_Y(t) = F'_Y(t) = [F_X(\gamma_0 t)]' = F'_X(\gamma_0 t)\gamma_0 = f_X(\gamma_0 t)\gamma_0, \tag{1.6}$$

$$h_Y(t) = \frac{f_Y(t)}{\bar{F}_Y(t)} = \frac{f_X(\gamma_0 t)\gamma_0}{\bar{F}_X(\gamma_0 t)} = h_X(\gamma_0 t)\gamma_0.$$

In the Two-Sample Accelerated Life Model (1.5), we also make the following observations. If $\gamma_0 > 1$, then the failure time of X is greater than the failure time of Y. In terms of the clinical study, this indicates that the treatment is effective. On the other hand, if $0 < \gamma_0 \le 1$, then the failure time of X is not greater than the failure time of Y, which indicates that the treatment is not effective.

In practice, the scale parameter γ_0 in (1.5) is unknown. For statistical inferences, we use the available data to estimate γ_0. To determine if the treatment is effective, the following hypothesis test may be considered:

$$H_0 : 0 < \gamma_0 \le 1 \quad \text{(treatment\ not\ effective)}$$
$$H_1 : \gamma_0 > 1 \qquad \text{(treatment\ effective).} \tag{1.7}$$

Also, point estimators and interval estimators for γ_0 based on the available data may be used to assess the effectiveness of the treatments.

1.2.2 General Accelerated Life Model

In the general case, we may be interested in the effects of several different variables or covariates on the failure time. For these cases, we consider, more generally, a covariate vector $\mathbf{z}$ of explanatory variables, and the general *Accelerated Life Model* is given by

$$T = \frac{T_0}{\psi(\mathbf{z})}, \tag{1.8}$$

where T_0 corresponds to the baseline lifetime random variable with $\mathbf{z} = \mathbf{0}$, and $\psi(\mathbf{z}) \geq 0$ is a function of the explanatory variables $\mathbf{z}$ satisfying $\psi(\mathbf{0}) = 1$. Below is an example on the general Accelerated Life Model (1.8).

Example 1. Consider a new treatment for lung cancer with $\mathbf{z} = (z_1, z_2, z_3)$, where $z_1 = $ gender; $z_2 = $ treatment; and $z_3 = $ [change in size of tumor]. In model (1.8), we have T as survival time from lung cancer treatment.

From model assumption (1.8), the following relationships among the distribution function, density function, and hazard function of the lifetime random variables T and T_0 are

implied. Let F_0 denote the distribution function of T_0, then we have

$$F(t; \mathbf{z}) \;=\; P\{T \le t\} \;=\; P\left\{ \frac{T_0}{\psi(\mathbf{z})} \le t \right\} \;=\; P\{T_0 \le t\psi(\mathbf{z})\} \;=\; F_0(t\psi(\mathbf{z})),$$

$$f(t; \mathbf{z}) \;=\; F'(t; \mathbf{z}) \;=\; F_0'(t\psi(\mathbf{z}))\psi(\mathbf{z}) \;=\; f_0(t\psi(\mathbf{z}))\psi(\mathbf{z}), \tag{1.9}$$

$$h(t; \mathbf{z}) \;=\; \frac{f(t; \mathbf{z})}{\bar{F}(t; \mathbf{z})} \;=\; \frac{f_0(t\psi(\mathbf{z}))\psi(\mathbf{z})}{\bar{F}_0(t\psi(\mathbf{z}))} \;=\; h_0(t\psi(\mathbf{z}))\psi(\mathbf{z}).$$

Moreover, taking the natural logarithm of both sides of (1.8) yields

$$\log T \;=\; \log T_0 - \log \psi(\mathbf{z}) \;=\; \mu_0 - \log \psi(\mathbf{z}) + \epsilon, \tag{1.10}$$

where μ_0 is the mean of $\log T_0$, and ϵ is a random variable with mean 0 whose distribution does not depend on $\mathbf{z}$.

In some cases, a parametric form for $\psi(\cdot)$ may be needed, say, in the form of $\psi(\mathbf{z}; \boldsymbol{\beta})$ by introducing a parameter $\boldsymbol{\beta}$. Since we require $\psi(\mathbf{z}) = \psi(\mathbf{z}; \boldsymbol{\beta}) \ge 0$ and $\psi(\mathbf{0}; \boldsymbol{\beta}) = \mathbf{1}$ in (1.10), a natural choice is

$$\psi(\mathbf{z}; \boldsymbol{\beta}) = e^{\boldsymbol{\beta}^T \mathbf{z}}. \tag{1.11}$$

With this choice, Accelerated Life Model (1.8) or (1.10) is written as

$$\log T = \mu_0 - \boldsymbol{\beta}^T \mathbf{z} + \epsilon, \tag{1.12}$$

which is the *log linear model* with parameter $\boldsymbol{\beta}$, and the usual linear model technique may used to study this model.

Two-Sample ALM vs. General ALM. In (1.11), we have $\psi(0, \beta) = 1$ and $\psi(1, \beta) = e^{\beta} \equiv \gamma_0$. Thus, a special case of (1.8) with function (1.11) for $z = 0, \; 1$ gives:

$$T = \frac{T_0}{\psi(0; \beta)} = T_0, \qquad T = \frac{T_0}{\psi(1; \beta)} = \frac{T_0}{\gamma_0}, \tag{1.13}$$

which coincides with the Two-Sample Accelerated Life Model (1.5). In the situation with a limited number of distinct values of $\mathbf{z}$, it is unnecessary to specify a parametric form for $\psi(\mathbf{z})$ such as (1.11), and with limited information, if any, about the underlying distribution, the choice of (1.11) may not even be reasonable. In cases where we have only a few treatment levels, we may use pairwise studies, i.e., the Two-Sample Accelerated Life Model (1.5), to compare the effects of different treatments. For these reasons, we focus on the Two-Sample Accelerated Life Model (1.5) throughout the rest of this dissertation.

Relation to Proportional Hazards Model. The Accelerated Life Model and the Cox Proportional Hazards model are the two main models used in survival analysis. The *Cox Proportional Hazards Model* (Cox, 1972) [8] is given by

$$h(t; \mathbf{z}) = h_0(t) e^{\boldsymbol{\beta}^\top \mathbf{z}}, \tag{1.14}$$

where $h(t; \mathbf{z})$ is the conditional hazard function of T given $\mathbf{Z} = \mathbf{z}$, and $h_0(t)$ is an arbitrary baseline (control group) hazard function. It has been shown that the Accelerated Life Model and Cox Proportional Hazards Model coincide if and only if the failure time follows a Weibull distribution (Cox and Oakes, 1984; page 71-72)[9]. When there is no evidence that the survival time follows a Weibull distribution or when the Cox Model assumption does not hold for the available data, it is vital to develop estimation and testing procedures for the Accelerated Life Model which is an important alternative model to the Cox Model in survival analysis.

Model Checking. The Accelerated Life Model is applicable when different levels of stress or different treatments are applied to subjects and each different level of stress or

treatment is in turn believed to increase or decrease the failure time of the subjects. As mentioned previously, we can study the effects of new medical treatments that are anticipated to increase the survival time of individuals in the study. Since the distributions of $\log T$ in (1.12) differ only by a translation for different values of $\mathbf{z}$, the variance of $\log T$ should be constant. A simple analysis where we calculate the mean and standard deviation at different treatment levels can help to determine if the model is adequate. In treatment levels where severe censoring is present, the mean and standard deviation may be over or underestimated. In these cases, we need to explore different methods to deal with the censoring issue, and it is always desirable to develop goodness-of-fit tests for assessment of the validity of the model assumptions. In the context of this research, goodness-of-fit tests for the Two-Sample Accelerated Life Model (1.5) with various types of censored data are important. If the data set does not fit the model assumption, then any statistical conclusion under model assumption (1.5) is not reliable. To our best knowledge, up to now there has been no work done on goodness-of-fit tests for the Two-Sample Accelerated Life Model (1.5) for complicated types of censored data such as doubly censored, interval censored, and partly interval censored data, which are described in Section 1.3.

1.2.3 Review of Recent Work

Rank-based monotone estimating functions are developed for the ALM with right censored observations in Jin, Lin, Wei and Ying (2003).[21] Using a resampling technique, which

does not involve nonparametric density estimation or numerical derivatives, they estimate the limiting covariance matrices. These estimators, which are given by the roots of non-monotone estimating equations based on the familiar weighted log-rank statistics, are shown to be consistent and asymptotically normal. These estimators can be obtained by linear programming, and two examples are provided to show that the proposed methods perform well in practical settings.

Using semiparametric transformation models, Cai and Cheng (2004) [5] analyze doubly censored data with additional assumptions on the right and left censoring variables. In their paper, inference procedures for the regression parameters are given, and the asymptotic distributions are studied.

Chen, Shen and Ying (2005) [7] also consider right censored data and propose a rank estimation procedure based on stratifying a Gehan-type extension of the Wilcoxon-Mann-Whitney estimating function. The resulting estimate is shown to be consistent and asymptotically normal. It is shown that the stratification poses little loss of information. These techniques can be done with linear programming.

Using the empirical likelihood method, Zhou (2005) [46] derives a test based on the rank estimators of the regression coefficient for the Accelerated Life Model with right censored data. Simulations and examples show that the chi-squared approximation to the distribution of the log empirical likelihood ratio performs well and has some advantages over the existing methods.

Odell, Anderson, and D'Agostino (1992) [30] study a Weibull-based accelerated failure time model with left and interval censored data. This approach assumes a parametric model. Maximum likelihood estimators (MLEs) are compared with Midpoint estimators (MDEs) and simulation studies indicate many instances where the MLE is superior to the MDE.

Betensky, Rabinowitz, and Tsiatis (2001) [4] study the accelerated failure time model with interval censored data using estimating equations computed using examination times from the same individual treated as if they had been obtained from different individuals. This approach does not involve computing the nonparametric maximum likelihood estimate of the distribution function. Simulation results are provided.

Komarek, Lesaffre, and Hilton (2005) [24] estimate parameters of an accelerated failure time model using a semiparametric approach they developed. In this approach, they use a P-spline smoothing technique which directly provides predictive survival distributions for fixed values of covariates with the presence of left, right, and interval censored data. Applications of this approach are provided as well.

Tian and Cai (2006) [41] use a novel approach to make inferences about the parameters in the accelerated failure time model for current status and interval censored data through an estimator constructed by inverting a Wald-type test for testing a null proportional hazards model. In addition, a Markov chain Monte Carlo based resampling method is proposed to obtain, simultaneously, the point estimator and a consistent estimator of its variance-covariance matrix. Extensive numerical studies are provided.

Komarek and Lesaffre (2008) [25] explore the relationship of covariates to the time to caries of permanent first molars. An accelerated failure time model with random effects is suggested, taking into account that the observations are clustered. These methods involve analyzing multivariate doubly and interval-censored data. Model parameters are estimated using a Bayesian approach with Markov chain Monte Carlo methodology.

To our best knowledge, up to now, the Accelerated Life Model has not been considered in literature, in a unified way, for all of the types of censored data considered in this dissertation.

1.3   Censored Data

Let

$$X_1, X_2, \ldots, X_n \tag{1.15}$$

be a random sample from an unknown distribution function F_0. In practice, we often do not actually observe this sample due to censoring. In the following subsections, we describe various types of censored data along with some real data examples, and summarize the asymptotic results on the nonparametric maximum likelihood estimator $\hat{F}_n$ for F_0.

1.3.1 Right Censored Data

The observed data for the random sample (1.15) are $\mathbf{O}_i = (V_i, \delta_i)$, $1 \le i \le n$, with

$$
V_i = \begin{cases} X_i & \text{if } X_i \le C_i, \ \delta_i = 1 \\[2ex] C_i & \text{if } X_i > C_i, \ \delta_i = 0, \end{cases}
\tag{1.16}
$$

where C_i is the *right censoring variable* and is independent of X_i. This type of censored data has been extensively studied in statistical literature in the past few decades.

   Data Example 1. *Heart Transplant Data.* In Miller and Halpern (1982), [28], a right censored data set is presented, and a brief description of the this data set is as follows. The Stanford heart transplantation program began in October 1967. By February 1980, 184 patients had received transplants. In this example, X_i is the survival time after a heart transplant. Of these 184 patients, 71 were still alive at the end of the study; thus for each of these patients, X_i occurs at some point after the study, resulting in 71 right censored observations.

   NPMLE and Asymptotic Properties: The likelihood function for F_0 based on right censored data (1.16) is given in Kaplan and Meier (1958). [22]. The *nonparametric maximum likelihood estimator* (NPMLE) $\hat{F}_n$ for F_0 is the function that maximizes this likelihood function. For right censored data, the product-limit estimator of Kaplan and Meier (1958) [22] is the unique NPMLE for F_0. For right censored data, Wellner (1982) [43] showed the asymptotic efficiency of the NPMLE $\hat{F}_n$. In Gill (1983) [14], it is shown that $\sqrt{n}(\hat{F}_n - F_0)$ weakly

converges to a centered Gaussian process under certain conditions. Also, it has been shown by Stute and Wang (1993) [40] that $||\hat{F}_n - F_0|| \xrightarrow{a.s.} 0$, as $n \to \infty$.

1.3.2 Doubly Censored Data

The observed data for the random sample (1.15) are $\mathbf{O}_i = (V_i, \delta_i)$, $1 \le i \le n$, with

$$V_i = \begin{cases} X_i & \text{if } D_i < X_i \le C_i, \ \delta_i = 1 \\ C_i & \text{if } X_i > C_i, \ \delta_i = 2 \\ D_i & \text{if } X_i \le D_i, \ \delta_i = 3, \end{cases} \tag{1.17}$$

where C_i is a *right censoring variable*, D_i is a *left censoring variable*, and (C_i, D_i) is independent of X_i with $P\{D_i < C_i\} = 1$. As mentioned previously in Subsection 1.2.3, the results in Cai and Cheng (2004) [5] only apply to a special case of above doubly censored data (1.17). Specifically, Cai and Cheng (2004) [5] impose the restrictive assumption that the left and right censoring variables are always known, but in (1.17) the censoring variables are not always observed and the left and right censoring variables C_i and D_i are never observed at the same time. Thus, up to now, there have been no works on the Accelerated Life Model (1.5) for doubly censored data (1.17).

Data Example 2. *African Infant Precocity.* A classic example of doubly censored data, discussed in Turnbull (1974) [42], comes from Leiderman *et al.* (1973) [26], and a brief description of this data set is as follows. In Leiderman *et al.* (1973) [26], a study was done in a community in Kenya to establish norms for infant development as compared to the known

standards in the United States and the United Kingdom. The data set contains information on 65 children born between July 1 and December 31, 1969. The infants were tested at approximately 2-month intervals, beginning in January, 1970, to see when they learned a certain task. In this example, X_i is the age of an infant when he/she can first perform the certain task. Some infants were able to perform the task at the first test; thus for these infants, X_i occurs at some point before the first test, resulting in left censored observations. On the other hand, some infants were never able to perform the task during the study; thus for these infants, X_i occurs at some point after the final test, resulting in right censored observations. For the remaining infants, the first time that they performed the task was observed, resulting in uncensored observations.

Data Example 3. *Effectiveness of Screening Mammograms.* In Ren and Peer (2000), [37] a doubly censored data set is studied, and a brief description of this data set is as follows. The study is based on the serial screening mammograms obtained in Njmegen, The Netherlands, 1981-1990. There were 289 patients in the study. In this example, X_i is the age at which the tumor can be detected for the ith patient when biennial mammographic screening is the only detection method. Of these patients, 45 had tumors observed at their first screening mammogram; thus for each of these patients, X_i occurs at some point before the study began, resulting in 45 left censored observations. On the other hand, 132 of the patients never had a tumor observed; thus for these patients, X_i occurs at some point after the study ended, resulting in 132 right censored observations. For the remaining 112 patients, a tumor was detected during the serial screening mammogram, i.e., the tumor was observed at one

mammogram at time $t = X_i$, but was not observed in the previous mammogram. Thus for each of these patients, X_i was actually observed, resulting in 112 uncensored observations.

NPMLE and Asymptotic Properties: The likelihood function for F_0 based on doubly censored data (1.17) is given in Mykland and Ren (1996). [29] The NPMLE $\hat{F}$ for F_0 based on the data (1.17) is the distribution function that maximizes the likelihood function. Mykland and Ren (1996) [29] give necessary and sufficient conditions for a self-consistent estimator for F_0 to be the NPMLE. In Turnbull (1974), [42] an iterative procedure was proposed to obtain an estimate for the survival function when the data is grouped. For the general case, Mykland and Ren (1996) [29] give an algorithm to compute the NPMLE $\hat{F}_n$. It has been shown by Chang and Yang (1987) [6] and Gu and Zhang (1993) [17] that $||\hat{F}_n - F_0|| \xrightarrow{a.s.} 0$, as $n \to \infty$. It is also shown in Gu and Zhang (1993) [17] that for doubly censored data (1.17), that $\sqrt{n}(\hat{F}_n - F_0)$ weakly converges to a centered Gaussian process under certain regularity conditions.

1.3.3 Interval Censored Data

Case 1: The observed data for the random sample (1.15) are $\mathbf{O}_i = (C_i, \delta_i)$, $1 \leq i \leq n$, with

$$\delta_i = I(X_i \leq C_i); \tag{1.18}$$

Case 2: The observed data for the random sample (1.15) are $\mathbf{O}_i = (C_i, D_i, \delta_i)$, $1 \le i \le n$, with

$$\delta_i = \begin{cases} 1 & \text{if } D_i < X_i \le C_i \\[2mm] 2 & \text{if } X_i > C_i, \\[2mm] 3 & \text{if } X_i \le D_i, \end{cases} \tag{1.19}$$

where C_i and D_i are independent of X_i and satisfy $P\{D_i < C_i\} = 1$.

Data Example 4. *HIV Data.* The following interval censored case 2 data (1.19) were encountered in AIDS reserach; see De Gruttola and Lagakos (1989),[10] and see Ren (2003) [34] for a detailed discussion, while a brief description is as follows.

In De Gruttola and Lagakos (1989), [10] an interval censored data set on

$$X = \{\text{time of HIV infection}\}$$

from AIDS research was presented. Since 1978, 262 people with Type A and B haemophilia have been treated at Hôpital Kremlin Bicêtre and Hôpital Cœur des Yvelines in France. For each individual, the only information available on X is $X \in [X_L, X_R]$, while it is assigned $X_L = 1$ if the individual was found to be infected with HIV on his/her first test for infection. Along with the retrospective tests for evidence of HIV infection, observations X_L and X_R were determined by the time at which the blood samples were stored. In this data set, time is measured in 6-month intervals, with $X = 1$ denoting July 1978, and one of the interests of the study is the distribution of X.

Kim, De Gruttola and Lagakos (1993) [23] gave the updated version of this data set for 104 individuals in the heavily treated group, i.e., patients who received at least 1000 μg/kg of blood factor for at least one year between 1982 and 1985. This data set always satisfies $X_L < X_R$, and it is associated with interval censored case 2 data (1.19) in the following way:

$$1 < X_L < X_R < \infty \quad \Longleftrightarrow \quad \delta = 1, D = X_L, C = X_R$$

$$1 < X_L < X_R = \infty \quad \Longleftrightarrow \quad \delta = 2, D = X_L, C = \infty$$

$$1 = X_L < X_R < \infty \quad \Longleftrightarrow \quad \delta = 3, D = 1, C = X_R.$$

Note that due to the way in which X_L and X_R were determined, we may assume that $[X_L, X_R]$ is independent of X, because the available blood samples were stored purely from haemophilia treatment which had nothing to do with HIV infection Thus, this is a real data example for interval censored case 2 data (1.19).

NPMLE and Asymptotic Properties: The likelihood functions for F_0 based on interval censored data (1.18) and (1.19) are given in Groeneboom and Wellner (1992), [15] respectively. The NPMLE $\hat{F}_n$ for F_0 in each case is the distribution function that maximizes the respective likelihood function. It is shown in Groeneboom and Wellner (1992) [15] that we have $||\hat{F}_n - F_0|| \xrightarrow{a.s.} 0$, as $n \to \infty$. For interval censored case 1 data (1.18), we have

$$n^{1/3}[\hat{F}_n(t_0) - F_0(t_0)] \xrightarrow{D} C_0 \mathbb{Z}, \text{ as } \ltimes \to \infty \tag{1.20}$$

where C_0 is a constant and $\mathbb{Z} = \operatorname{argmin}(W(T) + t^2)$ with W as the two sided Brownian motion starting from 0. For interval censored case 2 data (1.19), Wellner (1995) [44] and Groeneboom (1996) [16] showed that (1.20) holds under certain regularity conditions. But

the general convergence rate for $\hat{F}_n$ with interval censored case 2 data (1.19) is not known up to now.

1.3.4 Partly Interval Censored Data

Case 1: The observed data for the random sample (1.15) are

$$\mathbf{O}_i = \begin{cases} X_i & \text{if } 1 \leq i \leq k_0 \\ (C_i, \delta_i), & \text{if } k_0 + 1 \leq i \leq n, \end{cases} \qquad (1.21)$$

where $\delta_i = I\{X_i \leq C_i\}$ and C_i is independent of X_i.

General Case: The observed data for the random sample (1.15) are

$$\mathbf{O}_i = \begin{cases} X_i & \text{if } 1 \leq i \leq k_0 \\ (\mathbf{C}, \boldsymbol{\delta}_i), & \text{if } k_0 + 1 \leq i \leq n, \end{cases} \qquad (1.22)$$

where for N potential examination times $C_1 < \cdots < C_N$, letting $C_0 = 0$ and $C_{N+1} = \infty$, we have $\mathbf{C} = (C_1, \ldots, C_N)$ and $\boldsymbol{\delta}_i = (\delta_i^{(1)}, \ldots, \delta_i^{(N+1)})$ with $\delta_i^{(j)} = 1$, if $C_{j-1} < X_i \leq C_j$; 0, elsewhere. This means that for intervals $(0, C_1], (C_1, C_2], \ldots, (C_N, \infty)$, we know which one of them X_i falls into.

Data Example 5. *Framingham Heart Disease Study.* Odell *et al.* (1992) [30] discuss a partly interval censored data set originally found in Feinleib *et al.* (1975), [13] and a brief description of the data set is as follows. The original Framingham Heart Study began in 1949 to determine the genetic effects of risk factors. In the follow up study on the children of

the original patients we have 2,568 female children. Each of these individuals was observed at the Framingham Heart Study facilities in Boston, Massachusetts at three different exam times to determine the first occurrence of subcategory angina pectoris (AP) in coronary heart disease. In this example, X_i is the time that the ith patient acquires AP, and $C_1 < C_2 < C_3$ denote the three exam times. Also, $C_0 = 0$ and $C_4 = \infty$. For 8 of the patients, X_i is actually observed. None of the 2,568 patients had acquired AP by the first exam; thus X_i did not occur in the interval $[0, C_1)$ for any patients in the study. Of the 2,568 patients, 16 patients had not acquired AP by the second exam; thus for these patients, X_i occurs in the interval $[C_1, C_2)$. Of the 2,568 patients, 13 patients had not acquired AP by the second exam, but had acquired AP by the third exam; thus for these patients, X_i occurs in the interval $[C_2, C_3)$. The remaining 2,531 had not acquired AP by the third exam; thus for these patients, X_i occurs in the interval $[C_3, \infty)$, resulting in 2,531 right censored observations.

NPMLE and Asymptotic Properties: The likelihood functions for F_0 based on the partly interval censored data (1.21) and (1.22) are given in Huang (1999), [18] respectively. The NPMLE $\hat{F}_n$ for F_0 for each case is the distribution function maximizing the respective likelihood function. Huang (1999) [18] showed that for partly interval censored data (1.21) and (1.22), $||\hat{F}_n - F_0|| \xrightarrow{a.s.} 0$, as $n \to \infty$, and that $\sqrt{n}(\hat{F}_n - F_0)$ weakly converges to a centered Gaussian process under certain conditions.

1.4 Likelihood

In this section, we briefly review the likelihood methods. Specifically, Subsection 1.4.1 reviews parametric likelihood; Subsection 1.4.2 reviews empirical likelihood (Owen,1988); [31]; and Subsection 1.4.3 discusses weighted empirical likelihood (Ren, 2001). [33]

1.4.1 Parametric Likelihood

Consider a random sample $X_1, X_2, \ldots, X_n$ from a distribution with density function $f(x; \theta)$, where $\theta \in \mathbb{R}^{\shortparallel}$ is an unknown parameter. Heuristically, the likelihood function is the probability that we observe what we observed. This translates into the following *parametric likelihood function $L(\theta \mid \mathbf{X})$* for parameter θ:

$$L(\theta \mid \mathbf{X}) = P\{\text{Observe what we observed}\} = \prod_{i=1}^{n} f(X_i \mid \theta), \tag{1.23}$$

where $\mathbf{X} = (X_1, X_2, \ldots, X_n)$. If we can maximize $L(\theta \mid \mathbf{X})$ with respect to θ over the entire parameter space Θ, then the value $\hat{\theta}$, at which $L(\theta \mid \mathbf{X})$ attains its maximum, is called the *maximum likelihood estimator* (MLE) for θ.

Consider the following hypothesis test:

$$H_0 : \theta = \theta_0 \qquad \text{vs.} \qquad H_1 : \theta \neq \theta_0. \tag{1.24}$$

Then, the likelihood ratio test statistic is given by:

$$R(\mathbf{X}; \theta) \;=\; \frac{\sup_{H_0} L(\eta \mid \mathbf{X})}{\sup L(\eta; \mathbf{X})} \;=\; \sup_{\eta=\theta_0} \frac{L(\eta \mid \mathbf{X})}{L(\hat{\theta} \mid \mathbf{X})} \;=\; \frac{L(\theta_0 \mid \mathbf{X})}{L(\hat{\theta} \mid \mathbf{X})}, \tag{1.25}$$

and the rejection region for (1.24) is:

$$\{\mathbf{X} \mid R(\mathbf{X}; \theta) \leq c\} \tag{1.26}$$

for some predetermined constant $0 < c < 1$. With the level of significance $0 < \alpha < 1$, we can determine c in the following way:

$$\alpha = P\{\text{Type I error}\} = P\{\text{reject } H_0 \mid H_0\} = P\{R(\mathbf{X}; \theta) \leq c \mid \theta = \theta_0\}$$
$$= P\{R(\mathbf{X}; \theta_0) \leq c\} = P\{-2\log R(\mathbf{X}; \theta_0) \geq -2\log c\} \approx P\{\chi_1^2 \geq -2\log c\}, \tag{1.27}$$

because Wilks (1938) [45] showed that the limiting distribution of $-2\log R(\mathbf{X}; \theta_0)$ is a chi-squared distribution.

The acceptance region for (1.24) is

$$\{\mathbf{X} \mid R(\mathbf{X}; \theta) \geq c\}. \tag{1.28}$$

Let

$$\lambda(\eta) = \frac{L(\eta \mid \mathbf{X})}{L(\hat{\theta} \mid \mathbf{X})}. \tag{1.29}$$

Then, a $(1 - \alpha)100\%$ confidence interval for $\theta = \theta_0$ is given by:

$$C(\mathbf{X}) = \{\eta \mid \lambda(\eta) \geq c\}. \tag{1.30}$$

When using parametric likelihood methods based on (1.23), we assume that the data come from a known distribution up to an unknown parameter. The main problem with this method is that in practical situations, we may not know anything about the underlying distribution. Assuming an incorrect underlying distribution can lead to incorrect statistical conclusions. This is especially a concern in survival analysis where sample sizes are generally small or

22

moderate, thus usually there is no sufficient information to justify the parametric assumption on the underlying distribution. Hence, a nonparametric approach is not only desirable, but is essential in survival analysis. In the next subsection, we outline the nonparametric likelihood method, called the empirical likelihood method, which provides flexibility through the use of a likelihood function that requires no parametric assumption on the underlying distribution.

1.4.2 Empirical Likelihood

Consider a random sample $X_1, X_2, \ldots X_n$ from an unknown distribution function F_0. In Owen (1988) [31], the *empirical likelihood function* or *nonparametric likelihood function* is given as:

$$L(F) = \prod_{i=1}^{n} [F(X_i) - F(X_i-)], \tag{1.31}$$

where F is any distribution function. It is shown that the distribution function that maximizes (1.31) over all distribution functions F is the *empirical distribution function*, denoted by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I\{X_i \leq x\}, \quad -\infty < x < \infty. \tag{1.32}$$

Next, we review the empirical likelihood method which is analogous to the parametric likelihood method described in the previous section.

Assume that a parameter θ_{F_0} of F_0 can be expressed as $\theta_{F_0} = T(F_0)$, where $T(\cdot)$ is a statistical functional. Analogous to (1.24), we consider the following hypothesis test:

$$H_0 : \theta_{F_0} = \theta_0 \qquad \text{vs.} \qquad H_1 : \theta_{F_0} \neq \theta_0. \tag{1.33}$$

Then, the empirical likelihood ratio test statistic for (1.33) analogous to (1.25) is given by:

$$R(\mathbf{X} \,|\, F_0) \;=\; \frac{\sup_{F \in H_0} L(F)}{\sup_F L(F)} \;=\; \sup_{T(F)=\theta_0} \frac{L(F)}{L(F_n)}, \tag{1.34}$$

where F_n is given in (1.32) and the rejection region analogous to (1.26) is:

$$\{\mathbf{X} \mid R(\mathbf{X} \,|\, F_0) \le c\} \tag{1.35}$$

for some predetermined constant $0 < c < 1$. With the level of significance $0 < \alpha < 1$, analogous to (1.27), we can determine c in the following way:

$$\alpha \;=\; P\{\text{Type I error}\} \;=\; P\{\text{reject } H_0 \mid H_0\} \;=\; P\{R(\mathbf{X} \,|\, F_0) \le c \mid T(F_0) = \theta_0\}$$
$$= \; P\{R_0 \le c\} \;=\; P\{-2\log R_0 \ge -2\log c\} \;\approx\; P\{\chi_1^2 \ge -2\log c\}, \tag{1.36}$$

where under H_0 in (1.33)

$$R_0 = R(\mathbf{X} \,|\, F_0), \qquad \text{when } T(F_0) \;=\; \theta_0, \tag{1.37}$$

because Owen (1988) [31] showed that, usually, $-2\log R_0$ has a limiting chi-square distribution under the null hypothesis.

The acceptance region for (1.33) analogous to (1.28) is:

$$\{\mathbf{X} \mid R(\mathbf{X} \,|\, F_0) \ge c\}. \tag{1.38}$$

Similar to (1.25) and (1.29), let

$$\lambda(F) = \frac{L(F)}{L(F_n)}. \tag{1.39}$$

It can be shown that when $T(F)$ is continuous, a $(1-\alpha)100\%$ confidence region for $\theta_{F_0} = \theta_0$ analogous to (1.29)$-$(1.30) is given by

$$C(\mathbf{X}) \;=\; \left\{ \theta \;\middle|\; \sup_{T(F)=\theta} \lambda(F) \ge c \right\} \;=\; \{\theta = T(F) \mid \lambda(F) \ge c\}. \tag{1.40}$$

24

In the special case where the parameter of interest is the mean of F_0, i.e.,

$$\theta_{F_0} = T(F_0) = \int x dF_0(x), \tag{1.41}$$

Owen (1988) [31] showed that the confidence region (1.40) is an interval. Specifically, Owen (1988) [31] established the following theorem.

Theorem 1 (Owen, 1988) *Assume F_0 is non-degenerate with $\int |x|^3 \; dF_0 < \infty$. For $0 < c < 1$, and for $\theta_0 = \int x dF_0(x)$, we have in (1.40)*

$$C(\mathbf{X}) = [X_{L,n}, X_{U,n}], \tag{1.42}$$

where $X_{L,n} = \inf_F \int x \; dF$ and $X_{U,n} = \sup_F \int x \; dF$, and we have

$$\lim_{n\to\infty} P(X_{L,n} \leq \theta_0 \leq X_{U,n}) = \lim_{n\to\infty} P(-2\log R_0 \leq -2\log c) = P(\chi_1^2 \leq -2\log c), \tag{1.43}$$

where χ_1^2 is a random variable with a chi-squared distribution with 1 degree of freedom.

Generally, the empirical likelihood method approach is preferred over the parametric approach in areas such as survival analysis, where, as mentioned earlier, information is limited and we have no sufficient evidence to assume a parametric form for the underlying distribution. Another desirable property is that empirical likelihood based confidence intervals have been shown to have good coverage levels when compared to parametric approaches Owen (2001) [32]. Much work has been done using the methods of empirical likelihood. In particular, the NPMLE $\hat{F}_n$ for each type of censored data mentioned in Section 1.3 was

obtained through writing out the empirical likelihood function and maximizing it. For each type of censored data, the asymptotic properties of the NPMLE $\hat{F}_n$ have been studied and summarized in Section 1.3. However, one drawback of the empirical likelihood method is that usually it is difficult to incorporate a model assumption into the formulation of the likelihood function along with censored data. Recently, Ren (2001) [33] developed a new nonparametric method for censored data, called *weighted empirical likelihood*, which was successfully used to solve several difficult statistical inference problems with different types of censored data mentioned in Section 1.3. Based on these results, we use the weighted empirical likelihood method for the problem of Two Sample Accelerated Life Model (1.5) described in Sections 1.2−1.3. In the next subsection, we outline and discuss the weighted empirical likelihood method.

1.4.3   Weighted Empirical Likelihood

In Ren (2001) [33], the *weighted empirical likelihood function* is given in a simple form that is applicable to various types of censored data in a unified form. This simple form is convenient and more easily to be used for incorporating the model assumptions into the formulation of the likelihood function for censored data. Next, we describe the weighted empirical likelihood function and its applications.

As explained in Ren (2008a) [35], the weighted empirical likelihood method can be understood as follows. As in equation (1.15) from Section 1.3, we consider a random sample

$$X_1, \ldots, X_n \tag{1.44}$$

from an unknown distribution function F_0. Recall from Section 1.3, in practice we often do not observe the complete sample (1.44), instead we observe various types of censored data denoted by:

$$\mathbf{O}_1, \ldots, \mathbf{O}_n, \tag{1.45}$$

which is the observed censored sample for sample (1.44) and the data are possibly one of the types of censored data mentioned in Section 1.3; i.e., the $\mathbf{O}_i$'s in (1.45) could be right censored (1.16), doubly censored (1.17), interval censored Case 1 or Case 2 (1.18)$-$(1.19), or partly interval censored (1.21)$-$(1.22), etc. As reviewed in Section 1.3, the NPMLE $\hat{F}_n$ for F_0 has been studied for censored data (1.16)$-$(1.22), and it is shown that from the observed censored data (1.45), there exist m distinct points

$$W_1 < W_2 < \ldots < W_m, \tag{1.46}$$

along with $\hat{p}_j > 0, 1 \le j \le m$ such that the NPMLE $\hat{F}_n$ can be expressed as:

$$\hat{F}_n(x) = \sum_{i=1}^{m} \hat{p}_i I\{W_i \le x\}, \tag{1.47}$$

for right censored data (Kaplan and Meier, 1958), [22] doubly censored data (Mykland and Ren, 1996), [29] interval censored data Case 1 and Case 2 (Groenboom and Wellner, 1992), [15] and partly interval censored data (Huang, 1999). [18] Specifically, for right cen-

sored data (1.16), the W_i's are the noncensored observations and m is the number of un-censored observations (Kaplan and Meier, 1958). [22] For the more complicated types of censored data $(1.17)-(1.22)$ discussed in Section 1.3, the W_i's and $\hat{p}_i$'s are obtained through computing the NPMLE $\hat{F}_n$. Since, as reviewed in Section 1.3, the NPMLE $\hat{F}_n$ is shown to be a strong uniform consistent estimator for the underlying distribution F_0 under certain regu-larity conditions for the types of censored data aforementioned, we expect a random sample $X_1^*, \ldots, X_n^*$ taken from $\hat{F}_n$ to behave asymptotically the same as random sample (1.44). Let F_n^* denote the empirical distribution function (1.32) of the random sample $X_1^*, \ldots, X_n^*$, then we have $\hat{F}_n \approx F_n^*$, in turn, we have

$$
\begin{aligned}
\prod_{i=1}^{n} P\{X = X_i\} &\approx \prod_{i=1}^{n} P\{X^* = X_i^*\} \\
&= \prod_{j=1}^{m} \left(P\{X^* = W_j\} \right)^{n[F_n^*(W_j) - F_n^*(W_j-)]} \\
&\approx \prod_{j=1}^{m} \left(P\{X^* = W_j\} \right)^{n[\hat{F}_n(W_j) - \hat{F}_n(W_j-)]} \\
&= \prod_{j=1}^{m} \left(P\{X^* = W_j\} \right)^{n\hat{p}_j}.
\end{aligned}
\tag{1.48}
$$

Hence, the weighted empirical likelihood function for F_0 is given by

$$
\hat{L}(F) = \prod_{i=1}^{m} [F(W_i) - F(W_i-)]^{n\hat{p}_i},
\tag{1.49}
$$

where F is any distribution function and $\hat{F}_n$ maximizes $\hat{L}(F)$. Thus, the weighted empirical likelihood function $\hat{L}(F)$ may viewed as the asymptotic version of the empirical likelihood function $L(F)$ in (1.31) for censored data (Ren, 2008a). [35]

Note that when there is no censoring, it is shown (Ren, 2001) [33] that the weighted empir-ical likelihood function (1.49) coincides with the empirical likelihood function (1.31) given in

Owen (1988) .[31] Also, from the formulation of (1.49), the censoring mechanism is reflected in $\hat{L}(F)$ via the probability mass of the NPMLE $\hat{F}_n$ for F_0. Since the simple form of (1.49) depends only on the W_i's and $\hat{p}_i$'s obtained from the NPMLE $\hat{F}_n$, the weighted empirical likelihood method is easily applicable in a unified way to all of the types of censored data discussed in Section 1.3. In particular, once the W_i's and $\hat{p}_i$'s are computed from the NPMLE $\hat{F}_n$ with the specific type of censored data, the routines for computing weighted empirical likelihood based confidence intervals, test statistics, etc., are the same for the different types of censored data; thus weighted empirical likelihood simplifies the likelihood based computational problems for statistical inference problems with various types of censored data. Another advantage of the weighted empirical likelihood method based on (1.49) is that the theoretical and asymptotic results often can be obtained in a unified way via the statistical functional of the NPMLE $\hat{F}_n$ for different types of censored data. Moreover, the simple form of weighted empirical likelihood function (1.49) also makes it easier to incorporate model assumptions into the formulation of the likelihood function for complicated types of censored data, such as doubly censored data (1.17), interval censored Case 1 or Case 2 data $(1.18)-(1.19)$, and partly interval censored data $(1.21)-(1.22)$. For these complicated types of censored data, the resulting empirical likelihood function is usually very complicated and mathematically intractable.

It has been shown that the weighted empirical likelihood method can be used to solve difficult statistical inference problems with the various types of censored data mentioned above. For instance, in her recent work, Ren (2008a) [35] uses the weighted empirical likeli-

hood method to solve problems involving two sample semi-parametric models with various types of censored data, and Ren (2008b) [36] uses the weighted empirical likelihood method to construct smoothed weighted empirical likelihood ratio confidence intervals for quantiles with censored data. Based on the success of these recent works, since the problem of Two-Sample Accelerated Life Model (1.5) with complicated types of censored data described in Section 1.2, though very difficult on its own, is slightly related to the problems considered in Ren (2008a), [35] we apply the weighted empirical likelihood method to this problem in this research. It is through this weighted empirical likelihood method that we are able to provide solutions to the very difficult problems considered in this dissertation.

1.5   Centered Gaussian Processes

A Gaussian Process [39] $\mathbb{G}$ is a specific type of stochastic process such that the joint density function of any subset of the random variables is itself a multivariate Gaussian random variable. When each of these random variables have mean zero, we call this a centered Gaussian Process.

1.6   Summary of Main Results

In this research, we use the Weighted Empirical Likelihood approach to study the Accelerated Life Model with all of the types of censored data mentioned in Section 1.3. We outline

and compare inferences on the the scale parameter and the treatment using both Normal-Based Approximation and the weighted empirical likelihood approach. In particular, we obtain point estimates, hypothesis tests, and construct confidence intervals for both the scale parameter and the treatment mean. We also develop goodness-of-fit tests and provide simulation results to illustrate the effectiveness of our inferences.

This dissertation is organized as follows: Chapter 1 gives an introduction to the model and methods being utilized.

Chapter 2 introduces estimation for the Accelerated Life Model and discusses goodness of fit. In particular, in Section 2.2 we construct a Treatment Distribution Estimator for F_X in (1.4), we establish asymptotic properties for this estimator in Proposition 2.2 and establish rates of convergence for the estimator in Theorem 2.3 and Theorem 2.4 for right censored data (1.16), doubly censored data (1.17), and partly interval censored data (1.21)−(1.22). At the end of Section 2.3 we provide an approach for computing the p-value for a goodness of fit test statistic for right censored data (1.16), doubly censored data (1.17), and partly interval censored data (1.21)−(1.22).

Chapter 3 discusses estimation of the scale parameter in the Accelerated Life Model (1.5). In particular, in Section 3.1 we construct a naive estimator for γ_0 in (1.5), establish asymptotic properties for this estimator in Theorem 3.1 and construct normal based tests and confidence intervals for γ_0 based on this estimator. In Section 3.2 we discuss a rank based estimator and construct normal based tests and confidence intervals for γ_0 based on this estimator. We also provide an algorithm at the end of Section 3.2 for computing the rank

based estimator in practice which is applicable to all of the complicated types of censored data discussed in this dissertation. In Section 3.3 we discuss Weighted Empirical Likelihood Ratio based confidence intervals for γ_0 based on the rank based estimator.

Chapter 4 discusses estimation for the mean μ_X of the treatment group in (1.4). In particular, we construct point estimators for μ_X in Section 4.1. We construct normal based tests and confidence intervals for μ_X based on these point estimators in Section 4.2 and provide algorithms for computing these confidence intervals. In Section 4.3 we construct Weighted Empirical Likelihood Ratio based confidence intervals for μ_X based on the rank based point estimator and provide algorithms for computing these intervals in practice.

Chapter 5 discusses the bootstrap method and provides simulation results on the work described above.

Finally, Chapter 6 summarizes the research that has been done and provides direction for further development of the ideas in this dissertation.

CHAPTER 2
ESTIMATION AND GOODNESS OF FIT

In this chapter, we study the estimation problem for the treatment distribution function F_X in (1.5) and goodness-of-fit tests for the Two-Sample Accelerated Life Model (1.5). The methods developed in this chapter are applicable in a unified way to those different types of censored data described in Section 1.3. The organization of this chapter is as follows. Section 2.1 derives the weighted empirical likelihood function for (γ_0, F_X) under the Two-Sample Accelerated Life Model (1.5). Section 2.2 obtains an estimator for the treatment distribution function F_X in (1.5). Section 2.3 constructs goodness-of-fit tests for the Two-Sample Accelerated Life Model.

2.1   Weighted Empirical Likelihood for Accelerated Life Model

Consider the Two-Sample Accelerated Life Model (1.5), and consider that the two samples in (1.4) are censored data, denoted by:

$$\mathbf{O}_1^X, \ldots, \mathbf{O}_{n_1}^X \text{ is the observed sample for treatment sample } X_1, \ldots, X_{n_1}$$

$$\mathbf{O}_1^Y, \ldots, \mathbf{O}_{n_0}^Y \text{ is the observed sample for control sample } Y_1, \ldots, Y_{n_0},$$

(2.1)

where these two observed samples are independent, and the $\mathbf{O}_i^X$'s or $\mathbf{O}_i^Y$'s are possibly one of the types of censored data in Section 1.3, i.e., right censored data (1.16), doubly censored data (1.17), interval censored data Case 1 and Case 2 (1.18)$-$(1.19), or partly interval censored data (1.21)$-$(1.22). Note that it is not necessary for the two samples to be subject to the

same type of censoring; for instance, the data from the treatment group could be doubly censored and the data from the control group could be right censored.

Let $\hat{G}$ and $\hat{H}$ be the NPMLE for F_X and F_Y in (1.4) based on observed first and second censored data (2.1), respectively. As reviewed in Section 1.4.3, we know that there exist distinct points $W_1^X < \cdots < W_{m_1}^X$ and $W_1^Y < \cdots < W_{m_0}^Y$ as in (1.46) along with $\hat{p}_i^X > 0$ and $\hat{p}_j^Y > 0$ such that $\hat{G}$ and $\hat{H}$ can be expressed as

$$\hat{G}(x) = \sum_{i=1}^{m_1} \hat{p}_i^X I\{W_i^X \leq x\} \quad \text{and} \quad \hat{H}(x) = \sum_{j=1}^{m_0} \hat{p}_j^Y I\{W_j^Y \leq x\}, \tag{2.2}$$

respectively, for various types of censored data aforementioned. Note that $\hat{G}$ and $\hat{H}$ in (2.2) are not necessarily proper distribution functions. For this dissertation, we will adjust both $\hat{G}$ and $\hat{H}$ to proper distribution functions by setting $\hat{G} = 1$ and $\hat{H} = 1$ at the largest observations of the corresponding observed data sets in (2.1). Note that this adjustment implies that in (2.2) we have

$$\sum_{i=1}^{m_1} \hat{p}_i^X = 1 \quad \text{and} \quad \sum_{j=1}^{m_0} \hat{p}_j^Y = 1, \tag{2.3}$$

and that this kind of adjustment of the NPMLE is a generally adopted convention for censored data (Efron, 1967; Miller, 1976).[11] [27]

To derive the weighted empirical likelihood function for (γ_0, F_X) in Two-Sample Accelerated Life Model (1.5) based on observed two-sample censored data (2.1), we apply weighted empirical likelihood function (1.49) as follows. First, via $\hat{G}$ and $\hat{H}$ in (2.2), we apply weighted empirical likelihood function (1.49) to the two observed censored samples in (2.1), respectively. Since the two observed samples in (2.1) are independent, the weighted

empirical likelihood function based on the combined two samples in (2.1) is the product of the two weighted empirical likelihood functions. Thus, from the model assumption on the Two-Sample Accelerated Life Model (1.5)−(1.6) and from assumptions (2.2)−(2.3) for the NPMLE $\hat{G}$ and $\hat{H}$, we can write this weighted empirical likelihood function as follows:

$$\left(\prod_{i=1}^{m_1}[F_X(W_i^X) - F_X(W_i^X-)]^{n_1\hat{p}_i^X}\right)\left(\prod_{j=1}^{m_0}[F_Y(W_j^Y) - F_Y(W_j^Y-)]^{n_0\hat{p}_j^Y}\right)$$

$$\overset{(1.6)}{=} \left(\prod_{i=1}^{m_1}[F_X(W_i^X) - F_X(W_i^X-)]^{n_1\hat{p}_i^X}\right)\left(\prod_{j=1}^{m_0}\left[\gamma_0[F_X(\gamma_0 W_j^Y) - F_X(\gamma_0 W_j^Y-)]\right]^{n_0\hat{p}_j^Y}\right)$$

$$= \gamma_0^{n_0\sum_{j=1}^{m_0}\hat{p}_j^Y}\left(\prod_{i=1}^{m_1}[F_X(W_i^X) - F_X(W_i^X-)]^{n_1\hat{p}_i^X}\right)\left(\prod_{j=1}^{m_0}[F_X(\gamma_0 W_j^Y) - F_X(\gamma_0 W_j^Y-)]^{n_0\hat{p}_j^Y}\right)$$

$$\overset{(2.3)}{=} \gamma_0^{n_0}\left(\prod_{i=1}^{m_1}[F_X(W_i^X) - F_X(W_i^X-)]^{n_1\hat{p}_i^X}\right)\left(\prod_{j=1}^{m_0}[F_X(\gamma_0 W_j^Y) - F_X(\gamma_0 W_j^Y-)]^{n_0\hat{p}_j^Y}\right).$$

Hence, the weighted empirical likelihood function for (γ_0, F_X) in Two-Sample Accelerated Life Model (1.5) based on observed data (2.1) is given by

$$L(\gamma, F) = \gamma^{n_0}\left(\prod_{i=1}^{m_1}[F(W_i^X) - F(W_i^X-)]^{n_1\hat{p}_i^X}\right)\left(\prod_{j=1}^{m_0}[F(\gamma W_j^Y) - F(\gamma W_j^Y-)]^{n_0\hat{p}_j^Y}\right), \quad (2.4)$$

where F is any distribution function.

To simplify (2.4) for computational purpose, we introduce the following notations:

$$\mathbf{W}^{\boldsymbol{\gamma}} = (W_1^\gamma, \ldots, W_m^\gamma) = (W_1^X, \ldots, W_{m_1}^X, \gamma W_1^Y, \ldots, \gamma W_{m_0}^Y),$$

$$(w_1, \ldots, w_m) = (\rho_1\hat{p}_1^X, \ldots, \rho_1\hat{p}_{m_1}^X, \rho_0\hat{p}_1^Y, \ldots, \rho_0\hat{p}_{m_0}^Y), \quad (2.5)$$

where $m = m_0 + m_1$, $n = n_0 + n_1$, $\rho_1 = n_1/n$, and $\rho_0 = n_0/n$. From (2.5), weighted empirical likelihood function (2.4) can be rewritten as

$$L(\gamma, F) = \gamma^{n_0}\prod_{i=1}^{m} p_i^{nw_i}, \quad (2.6)$$

where γ is any positive real number, and F is given by

$$F(x) = \sum_{i=1}^{m} p_i I\{W_i^\gamma \leq x\}, \quad \text{for } p_i = F(W_i^\gamma) - F(W_i^\gamma-), \ 1 \leq i \leq m. \quad (2.7)$$

35

Hence, the weighted empirical likelihood based MLE $(\hat{\gamma}, \hat{F}_n)$ for (γ_0, F_X) is the solution that maximizes $L(\gamma, F)$ in (2.6).

2.2   Treatment Distribution Estimator

In this section, based on weighted empirical likelihood function (2.6) we derive an estimator for the treatment distribution F_X in Two-Sample Accelerated Life Model (1.5) with observed two-sample censored data (2.1), and we establish some asymptotic properties of this estimator for F_X. Note that the weighted empirical likelihood based MLE (WELMLE) $(\hat{\gamma}, \hat{F}_n)$ for (γ_0, F_X) is rather difficult to obtain, which requires additional restrictions and will be discussed later in Chapter 3. Here, we consider a simpler approach to obtain an estimator for F_X. The idea of this approach is that first, for a fixed $\gamma > 0$, we maximize $L(\gamma, F)$ over F, then we replace γ with a consistent estimator for γ_0.

For a fixed $\gamma > 0$, to maximize $L(\gamma, F)$ in (2.6) over all F given by (2.7), we need to solve the following optimization problem:

$$\begin{cases} \text{Maximize} \quad L(\gamma, F) = \gamma^{n_0} \prod_{i=1}^{m} p_i^{nw_i}, \\[2ex] \text{subject to:} \quad 0 \le p_i \le 1, \ 1 \le i \le m; \ \sum_{i=1}^{m} p_i = 1. \end{cases} \tag{2.8}$$

Note that if one of the p_i's above is 0, then $L(\gamma, F) = 0$. Also, note that if one of the p_i's is 1, then constraint $\sum_{i=1}^{m} p_i = 1$ implies that all other p_i's equal 0, thus $L(\gamma, F) = 0$. Hence,

optimization problem (2.8) is equivalent to

$$
\begin{cases}
\text{Maximize} \quad L(\gamma, F) = \gamma^{n_0} \prod_{i=1}^{m} p_i^{nw_i}, \\[2mm]
\text{subject to:} \quad 0 < p_i < 1, \ 1 \le i \le m; \ \sum_{i=1}^{m} p_i = 1.
\end{cases}
\tag{2.9}
$$

To solve optimization problem (2.9), we note that for all $1 \le i \le m$, we have

$$
\log L(\gamma, F) \ = \ n_0 \log \gamma + n \sum_{i=1}^{m} w_i \log p_i.
\tag{2.10}
$$

Thus, to find a candidate for the solution using the Lagrange Multipliers, we denote

$$
\mathscr{H}(\mathbf{p}, \lambda) = n \sum_{i=1}^{m} w_i \log p_i + \lambda \left[1 - \sum_{i=1}^{m} p_i \right],
\tag{2.11}
$$

then, we have for $1 \le i \le m$,

$$
0 = \frac{\partial \mathscr{H}}{\partial p_i} = \frac{nw_i}{p_i} - \lambda \quad \Rightarrow \quad p_i = nw_i/\lambda.
\tag{2.12}
$$

From (2.12) and constraint $\sum_{i=1}^{m} p_i = 1$ in (2.9), we have

$$
1 = \sum_{i=1}^{m} p_i = \sum_{i=1}^{m} \frac{nw_i}{\lambda},
\tag{2.13}
$$

which, along with (2.3) and (2.5), implies that

$$
\lambda = \sum_{i=1}^{m} nw_i = n \left[\sum_{i=1}^{m_1} w_i + \sum_{i=m_1+1}^{m} w_i \right] = n \left[\rho_1 \sum_{i=1}^{m_1} \hat{p}_i^X + \rho_0 \sum_{j=1}^{m_0} \hat{p}_j^Y \right]
$$
$$
= n[\rho_1 + \rho_0] = n[(n_1 + n_0)/n] = n_1 + n_0 = n.
\tag{2.14}
$$

Thus, from (2.12) and (2.14) a candidate for the solution of (2.9) is given by

$$
\hat{\mathbf{p}} = (w_1, \ldots, w_m).
\tag{2.15}
$$

The following lemma shows that this candidate is the unique solution for (2.9).

Lemma 2.1. *For any fixed $\gamma > 0$, $\hat{\mathbf{p}}$ in (2.15) is the unique solution of (2.9).*

Proof We prove that $\hat{\mathbf{p}} = (\hat{p}_1, \ldots, \hat{p}_m)$, given by (2.15), is the unique solution to (2.9) by verifying the *Karush-Kuhn-Tucker* (KKT) conditions in Theorem 4.3.8 of Bazarra, Sherali, and Shetty (1993; page 164) [1] as follows. From (2.11), let

$$h(\mathbf{p}) = n \sum_{i=1}^{m} w_i \log p_i \tag{2.16}$$

and let

$$A = \left\{ \mathbf{p} \,\middle|\, 0 < p_i < 1, \ 1 \leq i \leq m; \ \sum_{i=1}^{m} p_i = 1 \right\}. \tag{2.17}$$

The Hessian matrix (Bazarra, Sherali, and Shetty, 1993; page 90) [1] of $h(\mathbf{p})$ given in (2.16), exists on the set A and is given by

$$\frac{\partial^2 h(\mathbf{p})}{\partial p_i \partial p_j} = \begin{cases} -\dfrac{nw_i}{p_i^2} & \text{if } i = j \\[2ex] 0 & \text{if } i \neq j. \end{cases} \implies \mathbf{H}_h = \text{diag}\left\{ -\frac{nw_1}{p_1^2}, \ldots, -\frac{nw_m}{p_m^2} \right\}. \tag{2.18}$$

Since $\mathbf{H}_h$ is a diagonal matrix with diagonal elements $-\frac{nw_i}{p_i^2} < 0$, $1 \leq i \leq m$, for $\mathbf{p} \in A$, $\mathbf{H}_h$ is negative definite on A. Note that A is a convex set because for any $\mathbf{p}, \mathbf{q} \in A$ and $\mathbf{r} = \lambda \mathbf{p} + (1 - \lambda)\mathbf{q}$ with any $\lambda \in (0, 1)$ we have

$$0 < r_i = \lambda p_i + (1 - \lambda)q_i < \lambda + (1 - \lambda) = 1, \ 1 \leq i \leq m;$$
$$\sum_{i=1}^{m} r_i = \lambda \sum_{i=1}^{m} p_i + (1 - \lambda) \sum_{i=1}^{m} q_i = \lambda + (1 - \lambda) = 1. \tag{2.19}$$

Thus, function $h(\mathbf{p})$ is strictly concave on A by Theorem 3.3.8 of Bazarra, Sherali, and Shetty (1993; page 92 and page 79). [1] To verify the conditions in Theorem 4.3.8 of Bazarra, Sherali, and Shetty (1993; page 164), [1] note that $X_p = \{\mathbf{p} \mid 0 < p_i < 1, 1 \leq i \leq m\}$ is a nonempty open set in $\mathbb{R}^m$, and that $h(\mathbf{p})$ and $h_1(\mathbf{p}) = 1 - \sum_{i=1}^{m} p_i$ are both from $\mathbb{R}^{\geqslant} \to \mathbb{R}$. Since $\hat{\mathbf{p}} \in X_p$

38

satisfies constraint $h_1(\hat{\mathbf{p}}) = 0$, $\hat{\mathbf{p}}$ is a feasible solution for (2.9) (Bazarra, Sherali, and Shetty,

1993; page 99). [1] Also, note that with $v_1 = n$, the KKT conditions are satisfied because

$$\boldsymbol{\nabla} h(\hat{\mathbf{p}}) + v_1 \boldsymbol{\nabla} h_1(\hat{\mathbf{p}}) = \begin{pmatrix} nw_1/\hat{p}_1 \\ \vdots \\ nw_m/\hat{p}_m \end{pmatrix} + n \begin{pmatrix} -1 \\ \vdots \\ -1 \end{pmatrix} = \mathbf{0}.$$

Since $h(\mathbf{p})$ is concave and differentiable on A, $h(\mathbf{p})$ is psuedoconcave on A (Bazarra, Sherali,

and Shetty, 1993; page 116). [1] Since h_1 is a linear function, which means that h_1 is both

quasiconvex and quasiconcave on A (Bazarra, Sherali, and Shetty, 1993; pages 116, 118), [1]

by Theorems 3.4.2 and 4.3.8 of Bazarra, Sherali, and Shetty (1993; page 101, 164), [1] $\hat{\mathbf{p}}$ in

(2.15) is the unique global optimal solution to (2.9), which completes the proof.

To write the solution of (2.9) in form of (2.7) for any fixed $\gamma > 0$, we plug $\hat{\mathbf{p}}$ in (2.15)

into $F(x)$ in (2.7) and obtain:

$$
\begin{aligned}
\hat{F}_n(x; \gamma) &= \sum_{i=1}^{m} \hat{p}_i I\{W_i^\gamma \leq x\} \\
&= \sum_{i=1}^{m_1} w_i I\{W_i^\gamma \leq x\} + \sum_{i=m_1+1}^{m} w_i I\{W_i^\gamma \leq x\} \\
&= \rho_1 \sum_{i=1}^{m_1} \hat{p}_i^X I\{W_i^X \leq x\} + \rho_0 \sum_{j=1}^{m_0} \hat{p}_j^Y I\{\gamma W_j^Y \leq x\} \\
&= \rho_1 \sum_{i=1}^{m_1} \hat{p}_i^X I\{W_i^X \leq x\} + \rho_0 \sum_{j=1}^{m_0} \hat{p}_j^Y I\{W_j^Y \leq x/\gamma\} \\
&= \rho_1 \hat{G}(x) + \rho_0 \hat{H}(x/\gamma).
\end{aligned}
\tag{2.20}
$$

Thus, $\hat{F}_n(x; \gamma_0)$ is the WELMLE for treatment distribution F_X in (1.4)$-$(1.5) if γ_0 is known.

This means that in practice, if there exists a consistent estimator $\hat{\eta}$ for γ_0, an estimator for

F_X is given by

$$\hat{F}_n(x; \hat{\eta}) = \rho_1 \hat{G}(x) + \rho_0 \hat{H}(x/\hat{\eta}). \qquad (2.21)$$

The following proposition establishes some asymptotic properties on this estimator $\hat{F}_n(x; \hat{\eta})$.

Proposition 2.2. *Assume Two-Sample Accelerated Life Model (1.5) holds and assume*

(AS1) $\rho_0 = \frac{n_0}{n}$ *and* $\rho_1 = \frac{n_1}{n}$ *remain the same as* $n \to \infty$;

(AS2) $\sqrt{n} (\hat{\eta} - \gamma_0) \xrightarrow{D} N(0, \sigma_0^2)$, *as* $n \to \infty$;

(AS3) $\sqrt{n_1}(\hat{G} - F_X) \xrightarrow{w} \mathbb{G}_{\mathbb{X}}$, *as* $n_1 \to \infty$;

(AS4) $\sqrt{n_0}(\hat{H} - F_Y) \xrightarrow{w} \mathbb{G}_{\mathbb{Y}}$, *as* $n_0 \to \infty$;

where $\mathbb{G}_{\mathbb{X}}$ *and* $\mathbb{G}_{\mathbb{Y}}$ *are centered Gaussian processes. Then,* $\sqrt{n} (\hat{F}_n(\cdot; \hat{\eta}) - F_X)$ *weakly converges to a centered Gaussian process as* $n \to \infty$.

Remark 2.1. *On Assumptions of Proposition 2.2.* For assumption (AS2), it is not difficult to construct an estimator $\hat{\eta}$ for γ_0 satisfying (AS2), which will be studied for all the types of censored data (1.16)−(1.19) and (1.21)−(1.22) considered in this dissertation. For assumptions (AS3)−(AS4), as reviewed in Section 1.3, under certain regularity conditions these assumptions hold for right censored data (Gill, 1983) [14], doubly censored data (Gu and Zhang, 1993) [17], and partly interval censored data (Huang, 1999) [18]. For interval censored Case 1 or Case 2 data (1.18)−(1.19), assumptions (AS3)−(AS4) do not hold because, as reviewed in Section 1.3, Wellner (1995) [44] and Groenboom (1996) [16] showed that the convergence rate of the NPMLE for interval censored Case 1 data (1.18) is $n^{1/3}$, not $\sqrt{n}$; see (1.20) in Section 1.3.3. Up to now, the convergence rate of the NPMLE for interval censored Case 2 data (1.19) is not known.

Proof From (1.6), (2.5) and (2.21), we have:

$$\sqrt{n}\left[\hat{F}_n(x;\hat{\eta}) - F_X(x)\right] = \sqrt{n}\left[\rho_1\hat{G}(x) + \rho_0\hat{H}(x/\hat{\eta}) - F_X(x)\right]$$

$$= \sqrt{n}\left[\rho_1\hat{G}(x) - \rho_1 F_X(x) + \rho_0\hat{H}(x/\hat{\eta}) - \rho_0 F_X(x)\right]$$

$$= \sqrt{\frac{n_1}{\rho_1}}\rho_1\left[\hat{G}(x) - F_X(x)\right] + \sqrt{\frac{n_0}{\rho_0}}\rho_0\left[\hat{H}(x/\hat{\eta}) - F_X(x)\right]$$

$$= \sqrt{\rho_1}\sqrt{n_1}\left[\hat{G}(x) - F_X(x)\right] + \sqrt{\rho_0}\sqrt{n_0}\left[\hat{H}(x/\hat{\eta}) - F_Y(x/\gamma_0)\right]. \qquad (2.22)$$

Note that $\sqrt{\rho_1}\sqrt{n_1}\left[\hat{G}(x) - F_X(x)\right] \xrightarrow{w} \mathbb{G}_{\mathbb{X}}$. Considering the second part of 2.22, we have:

$$\sqrt{\rho_0}\sqrt{n_0}\left[\hat{H}(x/\hat{\eta}) - F_Y(x/\gamma_0)\right]$$

$$= \sqrt{\rho_0}\sqrt{n_0}\left[\hat{H}(x/\hat{\eta}) - F_Y(x/\hat{\eta}) + F_Y(x/\hat{\eta}) - F_Y(x/\gamma_0)\right]$$

$$= \sqrt{\rho_0}\sqrt{n_0}\left[\hat{H}(x/\hat{\eta}) - F_Y(x/\hat{\eta}))\right] + \sqrt{\rho_0}\sqrt{n_0}\left[F_Y(x/\hat{\eta}) - F_Y(x/\gamma_0)\right]$$

$$(2.23)$$

where $\sqrt{\rho_0}\sqrt{n_0}\left[\hat{H}(x/\hat{\eta}) - F_Y(x/\hat{\eta}))\right] \xrightarrow{w} G_Y(\hat{\eta})$ and $\sqrt{\rho_0}\sqrt{n_0}\left[F_Y(x/\hat{\eta}) - F_Y(x/\gamma_0)\right] \to \mathbb{Q}_{\mathbb{X}}$, as $n \to \infty$, where $\mathbb{Q}_{\mathbb{X}}$ is a centered Gaussian process. Thus, from $(AS1)-(AS4)$, $(2.22)-(2.23)$ and the fact that a linear combination of centered Gaussian processes is also a centered Gaussian process (Iranpour and Chacon, 1988; pg. 166), [20], we have that as $n \to \infty$, $\sqrt{n}\left[\hat{F}_n(x;\hat{\eta}) - F_X(x)\right]$ converges weakly to a centered Gaussian process, which completes the proof.

2.3 Goodness of Fit for Accelerated Life Model

Consider the following goodness-of-fit hypothesis test:

$$H_0 : \text{Two-Sample Accelerated Life Model (1.5) holds}$$

$$H_1 : \text{Two-Sample Accelerated Life Model (1.5) does not hold}$$

(2.24)

To construct a test statistic for (2.24), note that there are two ways to estimate F_X in Two-Sample Accelerated Life Model (1.5) based on censored data (2.1). One estimate is the NPMLE $\hat{G}$ for F_X given in (2.2), which is calculated using only the first sample in (2.1). Another estimate $\hat{F}_n(\,\cdot\,;\hat{\eta})$ given in (2.21) is calculated using two samples in (2.1) under H_0 in (2.24). Since $\hat{F}_n(\,\cdot\,;\hat{\eta})$ is derived under model assumption (1.5) and $\hat{G}$ is not, the difference between $\hat{F}_n(\,\cdot\,;\hat{\eta})$ and $\hat{G}$ measures the validity of the model assumption, and a large difference between $\hat{F}_n(\,\cdot\,;\hat{\eta})$ and $\hat{G}$ indicates that H_0 in (2.24) does not hold.

The following theorems establish the convergence rate for $\hat{F}_n(\,\cdot\,;\hat{\eta})$ for right censored data (1.16), doubly censored data (1.17), and partly interval censored data (1.21)$-$(1.22).

Theorem 2.3. *Under the assumptions of Proposition 2.2, we have $\sqrt{n}\left(\hat{F}_n(\,\cdot\,;\hat{\eta}) - \hat{G}\right)$ weakly converges to a centered Gaussian process as $n \to \infty$.*

Proof   From (2.5), we have:

$$\sqrt{n}\left[\hat{F}_n(x;\hat{\eta}) - \hat{G}(x)\right] = \sqrt{n}\left[\hat{F}_n(x;\hat{\eta}) - F_X(x) - \hat{G}(x) + F_X(x)\right]$$

$$= \sqrt{n}\left[\hat{F}_n(x;\hat{\eta}) - F_X(x)\right] - \sqrt{\frac{n}{n_1}}\sqrt{n_1}\left[\hat{G}(x) - F_X(x)\right]$$

$$= \sqrt{n}\left[\hat{F}_n(x;\hat{\eta}) - F_X(x)\right] - \frac{1}{\sqrt{\rho_1}}\sqrt{n_1}\left[\hat{G}(x) - F_X(x)\right].$$

Then, from Proposition 2.2 and assumptions (AS1) and (AS3) in Proposition 2.2, since the linear combination of centered Gaussian processes is also a centered Gaussian process (Iranpour and Chacon, 1988; pg. 166), [20] we have that as $n \to \infty$, $\sqrt{n} \left[\hat{F}_n(x; \hat{\eta}) - \hat{G}(x) \right]$ converges weakly to a centered Gaussian process, which completes the proof.

Theorem 2.4. *Assume* (AS1)$-$(AS4) *in Proposition 2.2. When Two-Sample Accelerated Life Model assumption (1.5) does not hold, we have* $\sqrt{n} \parallel \hat{F}_n(\cdot\,; \hat{\eta}) - \hat{G} \parallel \xrightarrow{P} \infty$, *as* $n \to \infty$.

Proof Assume that Two-Sample Accelerated Life Model (1.5) does not hold. From (2.5) and (2.21), we have:

$$
\begin{aligned}
\parallel \hat{F}_n(\cdot\,; \hat{\eta}) - \hat{G} \parallel &= \sup_{0 \le t < \infty} \left| \hat{F}_n(t; \hat{\eta}) - \hat{G}(t) \right| = \sup_{0 \le t < \infty} \left| \rho_1 \hat{G}(t) + \rho_0 \hat{H}(t/\hat{\eta}) - \hat{G}(t) \right| \\
&= \sup_{0 \le t < \infty} \left| \rho_0 \hat{H}(t/\hat{\eta}) + (\rho_1 - 1)\hat{G}(t) \right| = \sup_{0 \le t < \infty} \left| \rho_0 \hat{H}(t/\hat{\eta}) - \rho_0 \hat{G}(t) \right| \\
&= \rho_0 \sup_{0 \le t < \infty} \left| \hat{H}(t/\hat{\eta}) - \hat{G}(t) \right| = o_{a.s.}(1) + \rho_1 \| F_Y(t/\hat{\eta}) - F_X(t) \|
\end{aligned}
$$

We also have $\hat{\eta} \xrightarrow{P} \gamma_0$ which implies that $F_Y(t/\hat{\eta}) \to F_Y(t/\gamma_0)$ provided that d.f. F_Y is uniformly continuous. Note that if Two-Sample Accelerated Life Model (1.5) does not hold, from (1.6), we have $F_Y(t/\gamma_0) \ne F_X(t)$ and thus, $\parallel \hat{F}_n(\cdot\,; \hat{\eta}) - \hat{G} \parallel \to \delta > 0$, as $n \to \infty$, which completes this proof.

From Remark 1 and Theorems 2.3 and 2.4, for right censored data, doubly censored data, and partly interval censored data, we may use the following Kolmogorov-Smirnov type statistic (Serfling, 1980; page 63) [38] as the test statistic for goodness-of-fit test (2.24):

$$
T_n = \sqrt{n} \parallel \hat{F}_n(\cdot\,; \hat{\eta}) - \hat{G} \parallel = \sqrt{n} \sup_{0 \le t < \infty} \left| \hat{F}_n(t\,; \hat{\eta}) - \hat{G}(t) \right|. \tag{2.25}
$$

In practice, we need to compute the p-value for test statistic T_n in (2.25) based on observed data (2.1) for a given level of significance $0 < \alpha < 1$. One possible approach is given below.

Boostrap procedure for computing the p-value for test statistic T_n in (2.25):

Step 1. Generate bootstrap samples

$$\mathbf{O}_1^{X*}, \ldots, \mathbf{O}_{n_1}^{X*} \quad \text{and} \quad \mathbf{O}_1^{Y*}, \ldots, \mathbf{O}_{n_0}^{Y*}, \tag{2.26}$$

with replacement, from the two observed censored samples in (2.1), respectively.

Step 2. Compute $\hat{G}^*$ and $\hat{H}^*$ based on (2.2) and compute $\hat{\eta}^*$ using the bootstrap samples (2.26).

Step 3. Compute $\hat{F}_n^*(\,\cdot\,; \hat{\eta}^*)$ based on (2.21) as follows:

$$\hat{F}_n^*(x; \hat{\eta}^*) = \rho_1 \hat{G}^*(x) + \rho_0 \hat{H}^*(x/\hat{\eta}^*). \tag{2.27}$$

Step 4. Calculate the bootstrap estimate of test statistic (2.25) as follows:

$$T_n^* = \sqrt{n} \, \max_x \left| \left(\hat{F}_n^*(x; \hat{\eta}^*) - \hat{G}^*(x) \right) - \left(\hat{F}_n(x; \hat{\eta}) - \hat{G}(x) \right) \right|. \tag{2.28}$$

Step 5. Repeat Steps $1-4$ B times to obtain $T_n^*(b)$, $b = 1, \ldots B$, where B is usually chosen to be 1000. The bootstrap estimate for the p-value of goodness-of-fit test (2.24) is given by

$$\hat{p}^* = \frac{\#\{T_n^*(b) \geq T_n\}}{B}. \tag{2.29}$$

44

Remark 2.2. Note that if $\hat{\eta}$ is determined by $\hat{G}$ and $\hat{H}$ (see Chapter 3), i.e., $\hat{\eta} = \eta(\hat{G}, \hat{H})$, then $\hat{F}_n(t; \hat{\eta}) - \hat{G}(t)$ in (2.25) is a functional of $\hat{G}$ and $\hat{H}$ because

$$\hat{F}_n(t; \hat{\eta}) - \hat{G}(t) \stackrel{(2.21)}{=} \rho_1 \hat{G}(t) + \rho_0 \hat{H}(t/\eta(\hat{G}, \hat{H})) - \hat{G}(t) \equiv \tau(\hat{G}, \hat{H}). \qquad (2.30)$$

Thus, under Two-Sample Accelerated Life Model (1.5), assuming $\gamma_0 = \eta(F_X, F_Y)$, we have

$$\begin{aligned}
\tau(F_X, F_Y) &= \rho_1 F_X(t) + \rho_0 F_Y(t/\eta(F_X, F_Y)) - F_X(t) \\[2mm]
&= \rho_1 F_X(t) + \rho_0 F_Y(t/\gamma_0) - F_X(t) \\[2mm]
&\stackrel{(1.6)}{=} \rho_1 F_X(t) + \rho_0 F_X(t) - F_X(t) \stackrel{(2.5)}{\equiv} 0.
\end{aligned}$$

Hence, under model assumption (1.5), goodness-of-fit test statistic T_n in (2.25) can be expressed by:

$$T_n = \sqrt{n} \, \| \hat{F}_n(\cdot; \hat{\eta}) - \hat{G} \| = \sqrt{n} \, \| \tau(\hat{G}, \hat{H}) - \tau(F_X, F_Y) \| .$$

By the bootstrap principle given in Bickel and Ren (2001), [3] the distribution of T_n under H_0 in (2.24) can be consistently estimated by

$$\begin{aligned}
T_n^* &= \sqrt{n} \, \| \tau(\hat{G}^*, \hat{H}^*) - \tau(\hat{G}, \hat{H}) \| \\[2mm]
&\stackrel{(2.30)}{=} \sqrt{n} \, \max_x \left| \left(\rho_1 \hat{G}^*(x) + \rho_0 \hat{H}^*(x/\hat{\eta}^*) - \hat{G}^*(x) \right) - \left(\hat{F}_n(x; \hat{\eta}) - \hat{G}(x) \right) \right| \\[2mm]
&\stackrel{(2.27)}{=} \sqrt{n} \, \max_x \left| \left(\hat{F}_n^*(x; \hat{\eta}^*) - \hat{G}^*(x) \right) - \left(\hat{F}_n(x; \hat{\eta}) - \hat{G}(x) \right) \right| ,
\end{aligned}$$

where $\hat{\eta}^* = \eta(\hat{G}^*, \hat{H}^*)$, which gives (2.28). Such bootstrap method consistency relies on the n out of n bootstrap consistency for $\sqrt{n_1}(\hat{G} - F_X)$ estimated by $\sqrt{n_1}(\hat{G}^* - \hat{G})$ and for $\sqrt{n_0}(\hat{H} - F_Y)$ estimated by $\sqrt{n_0}(\hat{H}^* - \hat{H})$, respectively, which has been established for right censored data (1.16), doubly censored data (1.17), and partly interval censored

45

data $(1.21)-(1.22)$ by Bickel and Ren (1996) [2] and Huang (1999);[18] see the review in Section 1.3. Furthermore, it is worth noting that from Remark 1, we know that assumptions $(AS3)-(AS4)$ do not hold for interval censored Case 1 or Case 2 data $(1.18)-(1.19)$, but we still can check the validity of model assumption (1.5) graphically by comparing the curves of $\hat{F}_n(\,\cdot\,;\hat{\eta})$ and $\hat{G}$; i.e., if these curves differ obviously, then H_0 in (2.24) does not hold.

CHAPTER 3
ESTIMATION AND TESTS ON SCALE PARAMETER

In this chapter, we construct hypothesis tests and estimates for the scale parameter γ_0 in Two-Sample Accelerated Life Model (1.5) based on a naive estimator, a rank-based estimator and the Weighted Empirical Likelihood Method, respectively. The methods developed in this chapter are applicable in a unified way to those different types of censored data described in Section 1.3. The organization of this chapter is as follows. Section 3.1 gives a naive estimator $\hat{\gamma}_E$ for γ_0, based on which tests and confidence intervals are constructed. Section 3.2 considers a rank-based estimator $\hat{\gamma}_R$ for γ_0, based on which tests and confidence intervals are constructed. Section 3.3 constructs Weighted Empirical Likelihood Ratio based tests and confidence intervals for γ_0.

3.1   Naive Estimator

The idea of the construction of our naive estimator $\hat{\gamma}_E$ for γ_0 is as follows. Note that taking the expected value of both sides of equation (1.5), we obtain $E(Y) = E(X/\gamma_0)$, and by denoting $\mu_X = E(X)$ and $\mu_Y = E(Y)$, we obtain

$$\mu_Y = \frac{\mu_X}{\gamma_0} \quad \Longleftrightarrow \quad \gamma_0 = \frac{\mu_X}{\mu_Y}. \tag{3.1}$$

Thus, a *naive estimator* for γ_0 is naturally given by:

$$\hat{\gamma}_E \equiv \frac{\hat{\mu}_X}{\hat{\mu}_Y}, \tag{3.2}$$

where $\hat{\mu}_X$ and $\hat{\mu}_Y$ are estimators for μ_X and μ_Y, respectively; for instance, we have

$$\hat{\mu}_X = \int x \, d\hat{G}(x) = \sum_{i=1}^{m_1} \hat{p}_i^X W_i^X \quad \text{and} \quad \hat{\mu}_Y = \int x \, d\hat{H}(x) = \sum_{i=1}^{m_0} \hat{p}_i^Y W_i^Y \qquad (3.3)$$

with $\hat{G}$ and $\hat{H}$ given in (2.2). The following theorem establishes asymptotic properties of $\hat{\gamma}_E$ in (3.2).

Theorem 3.1. *Assume $\gamma_0 = \mu_X/\mu_Y$ and assume*

(AS5) $\quad \sqrt{n_1}(\hat{\mu}_X - \mu_X) \overset{D}{\to} N(0, \sigma_1^2)$, *as $n_1 \to \infty$*;

(AS6) $\quad \sqrt{n_0}(\hat{\mu}_Y - \mu_Y) \overset{D}{\to} N(0, \sigma_2^2)$, *as $n_0 \to \infty$.*

Then, $\sqrt{n}\,(\hat{\gamma}_E - \gamma_0) \overset{D}{\to} N(0, \sigma_E^2)$, as $n \to \infty$, where $\sigma_E^2 > 0$.

Remark 3.1. *Assumptions of Theorem 3.1.* Note that since the data in survival analysis are lifetimes, we always have $\mu_Y > 0$, hence $\gamma_0 = \mu_X/\mu_Y$ is well-defined. Also note that assumption $\gamma_0 = \mu_X/\mu_Y$ in Theorem 3.1 does not require Two-Sample Accelerated Life Model assumption (1.5). For assumption (AS5), note that as $n_1 \to \infty$,

$$
\begin{aligned}
\sqrt{n_1}(\hat{\mu}_X - \mu_X) &= \sqrt{n_1}\left[\int x \, d\hat{G}(x) - \int x \, dF_X(x)\right] \\
&= \sqrt{n_1}\left[\int (1 - \hat{G}(x)) \, dx - \int (1 - F_X(x)) \, dx\right] \\
&= -\int \sqrt{n_1}(\hat{G}(x) - F_X(x)) \, dx \overset{D}{\to} \int \mathbb{G}_0(x) \, dx, \qquad (3.4)
\end{aligned}
$$

because as reviewed in Section 1.3, $\sqrt{n_1}(\hat{G} - F_X)$ weakly converges to a centered Gaussian process $\mathbb{G}_0$ for right censored data (1.16), doubly censored data (1.17), and partly interval censored data (1.21)$-$(1.22). From properties of centered Gaussian processes (see Iranpour and Chacon (1988) pages 154-157 [20]), we know that $\int \mathbb{G}_0(x) \, dx$ is a zero-mean

normal random variable $N(0, \sigma_1^2)$ for some $0 < \sigma_1^2 < \infty$. Similarly, assumption (AS6) holds for the aforementioned types of censored data. For interval censored data Case 1 and Case 2 $(1.18)-(1.19)$, by a different argument, assumptions (AS5)$-$(AS6) also hold under certain conditions; see Huang and Wellner (1995) [19].

Proof From (2.5) and (3.2) we have

$$
\begin{aligned}
\sqrt{n}\left(\hat{\gamma}_E - \gamma_0\right) &= \sqrt{n}\left(\frac{\hat{\mu}_X}{\hat{\mu}_Y} - \frac{\mu_X}{\mu_Y}\right) = \sqrt{n}\left(\frac{\hat{\mu}_X \mu_Y - \hat{\mu}_Y \mu_X}{\hat{\mu}_Y \mu_Y}\right) \\
&= \sqrt{n}\left(\frac{\hat{\mu}_X \mu_Y - \mu_X \mu_Y + \mu_X \mu_Y - \hat{\mu}_Y \mu_X}{\hat{\mu}_Y \mu_Y}\right) \\
&= \sqrt{n}\left(\frac{\mu_Y(\hat{\mu}_X - \mu_X)}{\hat{\mu}_Y \mu_Y}\right) + \sqrt{n}\left(\frac{\mu_X(\mu_Y - \hat{\mu}_Y)}{\hat{\mu}_Y \mu_Y}\right) \\
&= \sqrt{n}\left(\frac{\hat{\mu}_X - \mu_X}{\hat{\mu}_Y}\right) + \sqrt{n}\left(\frac{\gamma_0(\mu_Y - \hat{\mu}_Y)}{\hat{\mu}_Y}\right) \\
&= \sqrt{\frac{n_1}{\rho_1}}\left(\frac{\hat{\mu}_X - \mu_X}{\hat{\mu}_Y}\right) + \sqrt{\frac{n_0}{\rho_0}}\left(\frac{\gamma_0(\mu_Y - \hat{\mu}_Y)}{\hat{\mu}_Y}\right) \\
&= \frac{1}{\hat{\mu}_Y}\left[\frac{1}{\sqrt{\rho_1}} \cdot \sqrt{n_1}\left(\hat{\mu}_X - \mu_X\right) + \frac{\gamma_0}{\sqrt{\rho_0}} \cdot \sqrt{n_0}\left(\mu_Y - \hat{\mu}_Y\right)\right] \\
&\xrightarrow{D} N(0, \sigma_E^2),
\end{aligned}
\tag{3.5}
$$

by (AS5) and (AS6), which completes the proof.

In the next two subsections, we construct tests and confidence intervals for γ_0 based on naive estimator $\hat{\gamma}_E$ given in (3.2).

3.1.1 Hypothesis Tests based on Normal Approximation

Recall from Section 1.2 that $\gamma_0 > 1$ in Two-Sample Accelerated Life Model (1.5) indicates that the treatment is effective. Hence, the following hypothesis test (which is analogous to

hypothesis test (1.7)) is often useful to assess the effectiveness of the treatment:

$$H_0 : \gamma_0 = 1 \quad \text{vs.} \quad H_1 : \gamma_0 > 1. \tag{3.6}$$

Based on point estimator $\hat{\gamma}_E$ for γ_0, in practice we reject H_0 in (3.6) if $\hat{\gamma}_E \geq c$ for some predetermined $c > 0$. For level of significance $0 < \alpha < 1$, we may determine c in practice via Theorem 3.1 as follows:

$$\begin{aligned} \alpha &= P\{\text{Type I Error}\} = P\{\text{reject } H_0 \mid H_0 \text{ is true}\} = P\{\hat{\gamma}_E \geq c \mid \gamma_0 = 1\} \\ &= P\left\{ \frac{\hat{\gamma}_E - \gamma_0}{\sigma_E/\sqrt{n}} \geq \frac{c - \gamma_0}{\sigma_E/\sqrt{n}} \;\middle|\; \gamma_0 = 1 \right\} \approx P\left\{ Z \geq \frac{c - 1}{\sigma_E/\sqrt{n}} \right\}, \end{aligned} \tag{3.7}$$

which gives

$$\frac{c - 1}{\sigma_E/\sqrt{n}} = z_\alpha \quad \Rightarrow \quad c = 1 + \frac{\sigma_E}{\sqrt{n}} z_\alpha, \tag{3.8}$$

where Z is the standard normal random variable, and z_α is the $(1 - \alpha)100$th percentile of Z.

In practice, we need to estimate the unknown parameter σ_E in (3.8). One possible approach is the following bootstrap procedure (Efron and Tibshirani, 1993) [12], which is valid for all of the types of censored data considered in this dissertation because of Theorem 3.1.

Boostrap procedure for estimating σ_E in (3.8)

Step 1. Generate bootstrap samples $\mathbf{O}_1^{X^*}, \ldots, \mathbf{O}_{n_1}^{X^*}$ and $\mathbf{O}_1^{Y^*}, \ldots, \mathbf{O}_{n_0}^{Y^*}$ as in (2.26).

Step 2. Compute $\hat{G}^*$ and $\hat{H}^*$ based on (2.2) using the bootstrap samples (2.26).

Step 3. Compute $\hat{\mu}_X^* = \int x \, d\hat{G}^*(x)$, $\hat{\mu}_Y^* = \int x \, d\hat{H}^*(x)$, and $\hat{\gamma}_E^* = \hat{\mu}_X^*/\hat{\mu}_Y^*$.

Step 4. Repeat Steps $1-3$ B times to obtain $\hat{\gamma}_E^*(b)$, $b = 1, \ldots B$, where B is usually chosen to be 1000. The bootstrap estimate for σ_E in (3.8) is given by

$$\hat{se}_{\hat{\gamma}_E} = \left(\sum_{b=1}^{B} \frac{\left(\hat{\gamma}_E^*(b) - \frac{1}{B} \sum_{i=1}^{B} \hat{\gamma}_E^*(i) \right)^2}{B-1} \right)^{1/2}. \tag{3.9}$$

3.1.2 Confidence Intervals based on Normal Approximation

A $(1 - \alpha)100\%$ confidence interval for γ_0 based on point estimator $\hat{\gamma}_E$ is constructed as follows. From Theorem 3.1, we have

$$
\begin{aligned}
1 - \alpha &= P\left\{ |Z| \leq z_{\alpha/2} \right\} \approx P\left\{ -z_{\alpha/2} \leq \frac{\hat{\gamma}_E - \gamma_0}{\sigma_E/\sqrt{n}} \leq z_{\alpha/2} \right\} \\
&= P\left\{ \hat{\gamma}_E - \frac{\sigma_E}{\sqrt{n}} z_{\alpha/2} \leq \gamma_0 \leq \hat{\gamma}_E + \frac{\sigma_E}{\sqrt{n}} z_{\alpha/2} \right\}.
\end{aligned} \tag{3.10}
$$

Thus, an approximated $(1 - \alpha)100\%$ confidence interval for γ_0 based on $\hat{\gamma}_E$ is given by

$$\hat{\gamma}_E \pm \frac{\sigma_E}{\sqrt{n}} z_{\alpha/2}, \tag{3.11}$$

where σ_E can be estimated by the above bootstrap procedure in (3.9), which gives

$$\hat{\gamma}_E \pm \frac{\hat{se}_{\hat{\gamma}_E}}{\sqrt{n}} z_{\alpha/2}. \tag{3.12}$$

Some simulation results on this are presented in Chapter 5.

3.2 Rank-Based Estimator

In this section, we discuss a rank-based estimator $\hat{\gamma}_R$ for γ_0 in Two-Sample Accelerated Life Model (1.5), construct hypothesis tests and confidence intervals analogous to those constructed in Section 3.1, and discuss computation of $\hat{\gamma}_R$.

A rank-based estimator for γ_0 is given as the solution of the following estimating equation:

$$2\sum_{i=1}^{m} \hat{G}(W_i^\gamma)p_i = 1, \tag{3.13}$$

where $\hat{G}$ is given in (2.2), and W_i^γ's and p_i's are given in (2.5)$-$(2.7). Since it is shown in Section 2.2 that $\hat{F}_n(\,\cdot\,;\gamma)$ given by (2.15)$-$(2.20) maximizes the weighted empirical likelihood function $L(\gamma, F)$ in (2.6) for any fixed γ, we plug $\hat{F}_n(\,\cdot\,;\gamma)$ into (3.13), i.e., $p_i = w_i$, to obtain the following estimating equation:

$$g(\gamma) = g(\gamma; \hat{G}, \hat{H}) \equiv \sum_{i=1}^{m} \left(\hat{G}(W_i^\gamma) - \frac{1}{2} \right) w_i \stackrel{\cdot}{\equiv} 0, \tag{3.14}$$

where "$\stackrel{\cdot}{\equiv}$" means that the solution of equation (3.14) is the value of γ where $g(\gamma)$ is closest to 0. Thus, our *rank-based estimator* $\hat{\gamma}_R$ for γ_0 is given by the solution of (3.14). Note that the use of "$\stackrel{\cdot}{\equiv}$" in (3.14) is necessary because it is shown later in Section 3.2.3 that $g(\gamma)$ is a piecewise step-function, hence equation $g(\gamma) = 0$ may not have an exact solution. Also in Section 3.2.3, we discuss additional properties of $g(\gamma)$ and computation of $\hat{\gamma}_R$.

Remark 3.2. *Rank-based estimating equations (3.13)$-$(3.14).* Note that the rank-based estimating equation given in (3.13)$-$(3.14) is applicable to all types of censored data considered in this dissertation, because $\hat{G}$ and $\hat{H}$ in (2.2), and W_i^γ's and w_i's in (2.5) are general notations for any types of censored data.

Ren (2008) [35] establishes the following theorem on asymptotic properties of $\hat{\gamma}_R$.

Theorem 3.2. *Assume (AS5)−(AS6). Then,* $\sqrt{n}\,(\hat{\gamma}_R - \gamma_0) \xrightarrow{D} N(0, \sigma_R^2)$, *as* $n \to \infty$, *where* $\sigma_R^2 > 0$.

In the next two subsections, we construct tests and confidence intervals for γ_0 based on rank-based estimator $\hat{\gamma}_R$.

3.2.1 Hypothesis Tests

Consider hypothesis test (3.6) from Section 3.1.1. Based on point estimator $\hat{\gamma}_R$ for γ_0, in practice we reject H_0 in (3.6) if $\hat{\gamma}_R \geq c$ for some predetermined $c > 0$. For level of significance $0 < \alpha < 1$, we may determine c in practice via Theorem 3.2 as follows:

$$
\begin{aligned}
\alpha &= P\{\text{Type I Error}\} = P\{\text{reject } H_0 \mid H_0 \text{ is true}\} = P\left\{\hat{\gamma}_R \geq c \mid \gamma_0 = 1\right\} \\
&= P\left\{\frac{\hat{\gamma}_R - \gamma_0}{\sigma_R/\sqrt{n}} \geq \frac{c - \gamma_0}{\sigma_R/\sqrt{n}} \,\bigg|\, \gamma_0 = 1\right\} \approx P\left\{Z \geq \frac{c - 1}{\sigma_R/\sqrt{n}}\right\},
\end{aligned}
\tag{3.15}
$$

which gives

$$
\frac{c - 1}{\sigma_R/\sqrt{n}} = z_\alpha \quad \Rightarrow \quad c = 1 + \frac{\sigma_R}{\sqrt{n}} z_\alpha.
\tag{3.16}
$$

In practice, we need to estimate the unknown parameter σ_R in (3.16). One possible approach is the following bootstrap procedure (Efron and Tibshirani, 1993) [12], which is valid for all of the types of censored data considered in this dissertation because of Theorem 3.2.

Boostrap procedure for estimating σ_R in (3.16)

Step 1. Generate bootstrap samples $\mathbf{O}_1^{X*}, \ldots, \mathbf{O}_{n_1}^{X*}$ and $\mathbf{O}_1^{Y*}, \ldots, \mathbf{O}_{n_0}^{Y*}$ as in (2.26).

Step 2. Compute $\hat{G}^*$ and $\hat{H}^*$ based on (2.2) using the bootstrap samples (2.26).

Step 3. Compute $\hat{\gamma}_R^*$, which is the solution of $g(\gamma; \hat{G}^*, \hat{H}^*) \doteq 0$ as in (3.14).

Step 4. Repeat Steps $1-3$ B times to obtain $\hat{\gamma}_R^*(b)$, $b = 1, \ldots B$, where B is usually chosen

to be 1000. The bootstrap estimate for the standard error is given by

$$\hat{se}_{\hat{\gamma}_R} = \left(\sum_{b=1}^{B} \frac{\left(\hat{\gamma}_R^*(b) - \frac{1}{B} \sum_{i=1}^{B} \hat{\gamma}_R^*(i) \right)^2}{B-1} \right)^{1/2}. \tag{3.17}$$

3.2.2 Confidence Intervals

A $(1-\alpha)100\%$ confidence interval for γ_0 based on point estimator $\hat{\gamma}_R$ is constructed as

follows. From Theorem 3.2, we have

$$\begin{aligned} 1 - \alpha &= P\left\{ |Z| \leq z_{\alpha/2} \right\} \approx P\left\{ -z_{\alpha/2} \leq \frac{\hat{\gamma}_R - \gamma_0}{\sigma_R/\sqrt{n}} \leq z_{\alpha/2} \right\} \\ &= P\left\{ \hat{\gamma}_R - \frac{\sigma_R}{\sqrt{n}} z_{\alpha/2} \leq \gamma_0 \leq \hat{\gamma}_R + \frac{\sigma_R}{\sqrt{n}} z_{\alpha/2} \right\}. \end{aligned} \tag{3.18}$$

Thus, an approximated $(1-\alpha)100\%$ confidence interval for γ_0 based on $\hat{\gamma}_R$ is given by

$$\hat{\gamma}_R \pm \frac{\sigma_R}{\sqrt{n}} z_{\alpha/2}, \tag{3.19}$$

where σ_R can be estimated by the above bootstrap procedure in (3.17), which gives

$$\hat{\gamma}_R \pm \frac{\hat{se}_{\hat{\gamma}_R}}{\sqrt{n}} z_{\alpha/2}. \tag{3.20}$$

Some simulation results on this are presented in Chapter 5.

3.2.3 Computability

In this subsection, we study the existence and uniqueness of the solution of $g(\gamma) \triangleq 0$ in (3.14), and provide an algorithm for computing $\hat{\gamma}_R$.

Some simulation results on $\hat{\gamma}_R$ are given in Chapter 5.

Existence and Uniqueness of Solution to $\boldsymbol{g(\gamma) \triangleq 0}$: To show the existence of a solution, we study the monotonic properties of $g(\gamma)$, and determine the behavior of $g(\gamma)$ at the end points of the interval $(0, \infty)$.

First, we simplify $g(\gamma)$ in (3.14) as follows:

$$
\begin{aligned}
g(\gamma) &= \sum_{i=1}^{m} \hat{G}(W_i^\gamma) w_i - \frac{1}{2} \\
&= \rho_1 \sum_{i=1}^{m_1} \hat{G}(W_i^X) \hat{p}_i^X + \rho_0 \sum_{i=1}^{m_0} \hat{G}(\gamma W_i^Y) \hat{p}_i^Y - \frac{1}{2} \\
&= \rho_1 \sum_{i=1}^{m_1} \left[\hat{p}_i^X \sum_{j=1}^{m_1} \hat{p}_j^X I\{W_j^X \le W_i^X\} \right] + \rho_0 \sum_{i=1}^{m_0} \left[\hat{p}_i^Y \sum_{j=1}^{m_1} \hat{p}_j^X I\{W_j^X \le \gamma W_i^Y\} \right] - \frac{1}{2} \\
&= \rho_1 \sum_{i=1}^{m_1} \left[\hat{p}_i^X \sum_{j=1}^{i} \hat{p}_j^X \right] + \rho_0 \sum_{i=1}^{m_0} \left[\hat{p}_i^Y \sum_{j=1}^{m_1} \hat{p}_j^X I\{W_j^X \le \gamma W_i^Y\} \right] - \frac{1}{2} \\
&= \rho_0 \sum_{i=1}^{m_0} \left[\hat{p}_i^Y \sum_{j=1}^{m_1} \hat{p}_j^X I\{(W_j^X / W_i^Y) \le \gamma\} \right] + Q \\
&= Q + \rho_0 \sum_{i=1}^{m_0} \sum_{j=1}^{m_1} \hat{p}_i^Y \hat{p}_j^X I\{(W_j^X / W_i^Y) \le \gamma\},
\end{aligned}
$$

(3.21)

(3.22)

(3.23)

where

$$
Q = \rho_1 \sum_{i=1}^{m_1} \left[\hat{p}_i^X \sum_{j=1}^{i} \hat{p}_j^X \right] - \frac{1}{2}. \tag{3.24}
$$

Note that Q is a constant (not depending on γ) and is calculated based on $\hat{G}$ in (2.2) via the treatment sample in (2.1). Thus, we have the following lemma on $g(\gamma)$ by (3.23).

Lemma 3.3. *The function $g(\gamma)$ in (3.14) has the following properties:*

(i) $g(\gamma)$ is a piecewise step-function;

(ii) $g(\gamma)$ is non-decreasing;

(iii) $g(\gamma)$ is right-continuous.

To study the values of $g(0)$ and $g(\infty)$, recall that, as reviewed in Section 1.3, for different types of censored data $(1.16)-(1.19)$ and $(1.21)-(1.22)$ considered in this dissertation, we know that under suitable conditions, we have $\| \hat{G} - F_X \| \xrightarrow{a.s.} 0$, as $n \to \infty$. Thus, since $\hat{G}(0) = 0$, in (3.21) we have $\hat{G}(\gamma W_i^Y) = 0$ for $\gamma = 0$ and

$$
\begin{aligned}
g(0) &= \rho_1 \int_0^\infty \hat{G}(x) \, d\hat{G}(x) - \frac{1}{2} = \rho_1 \left[\int_0^\infty F_X(x) \, dF_X(x) + o_p(1) \right] - \frac{1}{2} \\
&= \rho_1 \int_0^1 x \, dx + o_p(1) - \frac{1}{2} = \frac{\rho_1 - 1}{2} + o_p(1),
\end{aligned}
$$

where $o_p(1)$ converges to 0 in probability as $n \to \infty$. In turn, we have that as $n \to \infty$,

$$g(0) < 0 \qquad \text{in probability,} \tag{3.25}$$

because $0 < \rho_1 < 1$. Similarly, since $\hat{G}(\infty) = 1$, in (3.21) we have $\hat{G}(\gamma W_i^Y) = 1$ for $\gamma = \infty$ and by (2.3),

$$
\begin{aligned}
g(\infty) &= \rho_1 \int_0^\infty \hat{G}(x) \, d\hat{G}(x) + \rho_0 - \frac{1}{2} = \rho_1 \left[\int_0^\infty F_X(x) \, dF_X(x) + o_p(1) \right] + \rho_0 - \frac{1}{2} \\
&= \rho_1 \int_0^1 x \, dx + \rho_0 - \frac{1}{2} + o_p(1) = \frac{\rho_1}{2} + \rho_0 - \frac{1}{2} + o_p(1) = \frac{1 - \rho_1}{2} + o_p(1).
\end{aligned}
$$

Thus, we have that as $n \to \infty$,

$$g(\infty) > 0 \qquad \text{in probability.} \tag{3.26}$$

Note that by definition, $g(\gamma) \doteqdot 0$ in (3.14) always has a solution. But, by Lemma 3.3 (i), $g(\gamma) = 0$ does not necessarily have a solution. However, Lemma 3.3 and (3.25)−(3.26) imply that in probability, all scenarios for $g(\gamma)$ are as shown in Figures 3.1−3.3.
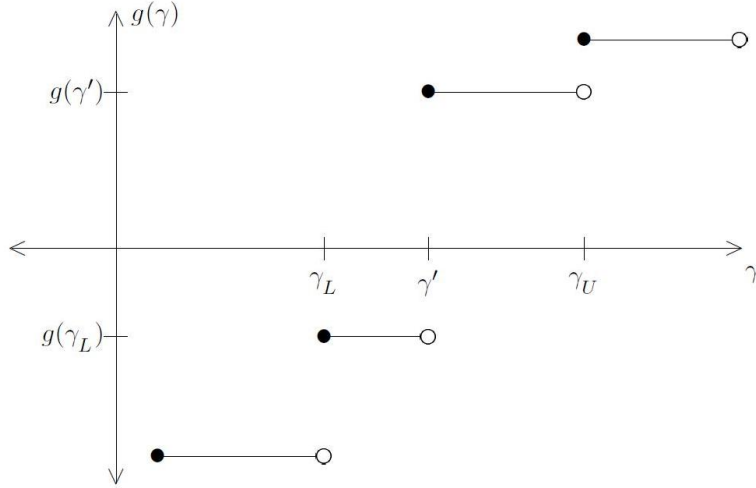
Figure 3.1: Scenarios for $g(\lambda)$: $|g(\gamma_L)| < |g(\gamma')|$
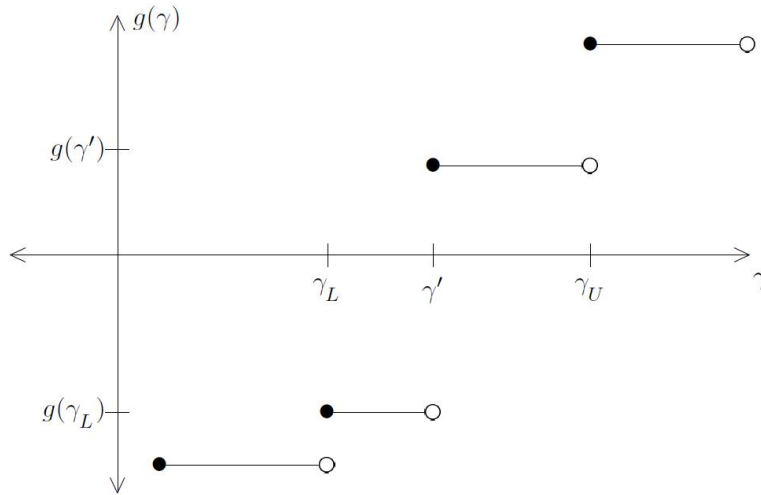
Figure 3.2: Scenarios for $g(\lambda)$: $|g(\gamma_L)| > |g(\gamma')|$
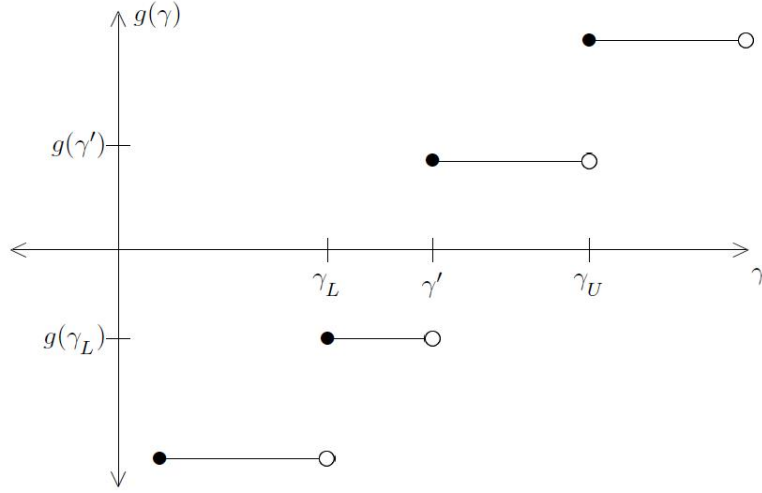
Figure 3.3: Scenarios for $g(\lambda)$: $|g(\gamma_L)| = |g(\gamma')|$

From Figures 3.1−3.3, clearly we see that there exists a unique value γ' such that in probability

$$g(\gamma'-) < 0 \qquad \text{and} \qquad g(\gamma') \geq 0, \tag{3.27}$$

which means that the solution of $g(\gamma) \doteq 0$ is not unique.

Throughout this dissertation, we define the solution $\hat{\gamma}_R$ of (3.14) as follows:

$$\hat{\gamma}_R = \begin{cases} \dfrac{\gamma_L + \gamma'}{2} & \text{if } |g(\gamma_L)| < |g(\gamma')| \\[2mm] \dfrac{\gamma' + \gamma_U}{2} & \text{if } |g(\gamma_L)| > |g(\gamma')| \\[2mm] \dfrac{\gamma_L + \gamma_U}{2} & \text{if } |g(\gamma_L)| = |g(\gamma')|, \end{cases} \tag{3.28}$$

where

$$\gamma_L = \inf \left\{ \gamma \mid g(\gamma) = g(\gamma'-) \right\} \qquad \text{and} \qquad \gamma_U = \sup \left\{ \gamma \mid g(\gamma) = g(\gamma') \right\}. \tag{3.29}$$

58

Algorithm for Computing $\hat{\boldsymbol{\gamma}}_{\boldsymbol{R}}$: In order to compute $\hat{\gamma}_R$ in (3.28), we need to find γ' in (3.27). Note that from Lemma 3.3 (ii) and (3.25)−(3.26), γ' can be found using a bisection algorithm, but the resulting solution is an estimated one. On the other hand, from Lemma 3.3 (i)−(ii) and (3.23), we know that $g(\gamma)$ only has jumps at the following points:

$$S_{JP} = \left\{ \frac{W_i^X}{W_j^Y} \,\middle|\, i = 1, \ldots, m_1, \; j = 1, \ldots, m_0 \right\}. \tag{3.30}$$

Thus, let

$$U_1 < U_2 < \ldots < U_{N-1} < U_N \tag{3.31}$$

denote all distinct points in S_{JP} in ascending order, then from Figures 3.1−3.3, we know that γ' is one of the values among $U_1, \ldots, U_N$. Hence, to find γ', we need to find point $U_{N'} = \gamma'$ such that (3.27) holds. The following procedure outlines how to compute $\hat{\gamma}_R$ in practice:

Bootstrap procedure for computing $\hat{\gamma}_R$:

Step 1. Obtain the values $U_1 < U_2 < \ldots < U_N$ in (3.31);

Step 2. Find $U_{N'}$ such that (3.27) holds for $U_{N'} = \gamma'$;

Step 3. Compute $\hat{\gamma}_R$ by (3.28)−(3.86).

Remark 3.3. Note that this algorithm for computing $\hat{\gamma}_R$ is applicable to all types of censored data considered in this dissertation, because rank-based estimating equation (3.14) is applicable to these different types of censored data; see Remark 3.2. Also, note that the computational efficiency in Step 2 is essential in this algorithm. While there are many different methods for finding γ', the approach we used in this dissertation is based on the idea of "bisection" and it performs well.

3.3 Weighted Empirical Likelihood Ratio Tests and Confidence Intervals

Recall that in Section 2.2, it is mentioned that additional restrictions are required to obtain the WELMLE $(\hat{\gamma}, \hat{F}_n)$ for (γ_0, F_X) in Two-Sample Accelerated Life Model (1.5). To see this, consider the following natural optimization problem for the weighted empirical likelihood function $L(\gamma, F)$ given by (2.6):

$$
\begin{cases}
\text{Maximize} \quad L(\gamma, F) \;=\; \gamma^{n_0} \prod_{i=1}^{m} p_i^{nw_i} \\
\text{subject to:} \quad 0 \le p_i \le 1, \; 1 \le i \le m; \; \sum_{i=1}^{m} p_i = 1; \; 0 < \gamma < \infty.
\end{cases}
\tag{3.32}
$$

For $F(x)$ in (2.7), let $F_m(x) = \sum_{i=1}^{m} \frac{1}{m} I\{W_i^{\gamma} \le x\}$. Then, we have in (3.32):

$$
L(\gamma, F_m) = \gamma^{n_0} \left(\frac{1}{m} \right)^{mnw_i} \to \infty, \qquad \text{as } \gamma \to \infty,
$$

which means that the solution of optimization problem (3.32) is ∞. This suggests that to have a finite solution in (3.32), we need additional constraint on γ. So far in statistical literature, Zhou (2005) [46] dealt with this issue by using a rank-based estimating equation on γ for right censored data (1.16), which came from Jin, Lin, Wei and Ying (2003) [21]. However, Zhou's estimating equation is not obviously applicable to the complicated types of censored data considered in this dissertation. One possible constraint on γ in (3.32) is to let $\gamma = \hat{\gamma}_E$, where $\hat{\gamma}_E$ given in (3.2) is applicable to all types of censored data considered in this dissertation. However, our studies show that this constraint results in a trivial solution and leads to nowhere. Another possible constraint on γ in (3.32) is to let γ be the solution to rank-based estimating equation (3.13), which, as discussed in Remark 3.2, is applicable to

all types of censored data considered in this dissertation. The optimization problem (3.32) with this rank-based estimating equation (3.13) is written as follows:

$$
\begin{cases}
\text{Maximize} \quad L(\gamma, F) \;=\; \gamma^{n_0} \prod_{i=1}^{m} p_i^{n w_i} \\[2ex]
\text{subject to:} \;\; 0 \le p_i \le 1, \; 1 \le i \le m; \; \sum_{i=1}^{m} p_i = 1; \; 2 \sum_{i=1}^{m} \hat{G}(W_i^{\gamma}) p_i = 1.
\end{cases}
\tag{3.33}
$$

By the same argument used in $(2.8)-(2.9)$, we know that for F given in (2.7), optimization problem (3.33) is equivalent to:

$$
\begin{cases}
\text{Maximize} \quad L(\gamma, F) \;=\; \gamma^{n_0} \prod_{i=1}^{m} p_i^{n w_i} \\[2ex]
\text{subject to:} \;\; 0 < p_i < 1, \; 1 \le i \le m; \; \sum_{i=1}^{m} p_i = 1; \; 2 \sum_{i=1}^{m} \hat{G}(W_i^{\gamma}) p_i = 1.
\end{cases}
\tag{3.34}
$$

In the next two subsections, via (3.34) we construct Weighted Empirical Likelihood Ratio based tests and confidence intervals for γ_0.

3.3.1 Hypothesis Tests

Under Two-Sample Accelerated Life Model (1.5), we consider the following hypothesis test for a given value $\gamma_{00} > 0$, which is more general than that in (3.6):

$$
H_0 : \gamma_0 = \gamma_{00} \quad \text{vs.} \quad H_1 : \gamma_0 \ne \gamma_{00}.
\tag{3.35}
$$

Under (3.34), the weighted empirical likelihood ratio test statistic for (3.35) analogous to (1.34) is given by:

$$
R(\mathbf{O}_n) \;=\; \frac{\sup_{(\gamma, F) \in H_0} L(\gamma, F)}{\sup_{(\gamma, F)} L(\gamma, F)} \;=\; \sup_{(\gamma, F) \in H_0} \frac{L(\gamma, F)}{L(\hat{\gamma}, \hat{F}_n)},
\tag{3.36}
$$

where the supremum is taken over (γ, F) satisfying the constraints in (3.34), $(\hat{\gamma}, \hat{F}_n)$ is the solution of (3.34), and by (2.1),

$$\mathbf{O}_n = \left\{ \mathbf{O}_1^X, \ldots, \mathbf{O}_{n_1}^X, \mathbf{O}_1^Y, \ldots, \mathbf{O}_{n_0}^Y \right\}. \tag{3.37}$$

To compute (3.36), we consider the computation of the denominator and numerator of (3.36), respectively, as follows.

Computation of the denominator in (3.36):

To solve (3.34), note that for any fixed $\gamma > 0$, $\hat{F}_n(\,\cdot\,; \gamma)$ given by (2.15)$-$(2.20) maximizes $L(\gamma, F)$ over all F satisfying the constraints in (2.9); see Section 2.2. Plugging $\hat{F}_n(\,\cdot\,; \gamma)$, i.e., $p_i = w_i$, into the last constraint equation in (3.34), we obtain the following equation:

$$2 \sum_{i=1}^{m} \hat{G}(W_i^\gamma) w_i = 1 \iff g(\gamma) = 0, \tag{3.38}$$

where $g(\gamma)$ is given in (3.14). From Section 3.2, we know that $g(\gamma) = 0$ does not necessarily have a solution. Thus, for solution $\hat{\gamma}_R$ of $g(\gamma) \doteqdot 0$ as in (3.14), the approximated solution $(\hat{\gamma}, \hat{F}_n)$ to optimization problem (3.34) is given by:

$$\hat{\gamma} = \hat{\gamma}_R \quad \text{and} \quad \hat{F}_n = \hat{F}_n(\,\cdot\,; \hat{\gamma}_R). \tag{3.39}$$

From (2.5)$-$(2.6), (2.15), (3.34) and (3.39), the denominator of (3.36) is given by

$$L(\hat{\gamma}, \hat{F}_n) = \hat{\gamma}^{n_0} \prod_{i=1}^{m} w_i^{n w_i}. \tag{3.40}$$

Note that for the rest of this dissertation, we treat $(\hat{\gamma}, \hat{F}_n)$ in (3.39) as the WELMLE for (γ_0, F_X) in Two-Sample Accelerated Life Model (1.5).

Computation of the numerator in (3.36): Note that the numerator is given by the solution of (3.34) under H_0 in (3.35), i.e., we need to find the solution of

$$\sup\left\{L(\gamma_{00}, F) \ \Big| \ 0 < p_i < 1, 1 \le i \le m; \ \sum_{i=1}^{m} p_i = 1; \ 2\sum_{i=1}^{m} \hat{G}(W_i)p_i = 1\right\}, \qquad (3.41)$$

where p_i's are given by (2.7) with $\gamma = \gamma_{00}$, and from (2.5) we use the following notation:

$$(W_1, \ldots, W_m) \ = \ \mathbf{W}^{\gamma_{00}} \ = \ (W_1^{\gamma_{00}}, \ldots, W_m^{\gamma_{00}}). \qquad (3.42)$$

To solve optimization problem (3.41), we note that for all $0 < p_i < 1$, we have

$$\log L(\gamma_{00}, F) \ = \ n_0 \log \gamma_{00} + n\sum_{i=1}^{m} w_i \log p_i.$$

Thus, to find a candidate for the solution using the Lagrange Multipliers, we denote

$$\mathscr{H}(\mathbf{p}, \beta, \lambda) = \sum_{=} \log_i + \beta\left[-\sum_{=} {}_i\right] + \lambda\left[-\sum_{=} \hat{\mathscr{G}}(\mathscr{W})_i\right], \qquad (3.43)$$

then, we have for $1 \le i \le m$,

$$0 = \frac{\partial \mathscr{H}}{\partial p_i} = \frac{nw_i}{p_i} - \beta - 2n\lambda\hat{G}(W_i) \quad \Rightarrow \quad p_i = \frac{nw_i}{\beta + 2n\lambda\hat{G}(W_i)}. \qquad (3.44)$$

From (2.3), (2.5), (3.44) and the last two constraints in (3.41), we have

$$\beta p_i = nw_i - 2n\lambda\hat{G}(W_i)p_i \quad \Rightarrow \quad \beta\sum_{i=1}^{m} p_i = n\sum_{i=1}^{m} w_i - 2n\lambda\sum_{i=1}^{m} \hat{G}(W_i)p_i$$
$$\Rightarrow \quad \beta = n - n \quad \Rightarrow \quad \beta = n(1 - \lambda). \qquad (3.45)$$

Plugging β in (3.45) into (3.44), we have

$$p_i = \frac{nw_i}{n(1 - \lambda) + 2n\lambda\hat{G}(W_i)} = \frac{w_i}{1 + \lambda(2\hat{G}(W_i) - 1)}, \qquad 1 \le i \le m. \qquad (3.46)$$

63

From the constraints in (3.41) and p_i in (3.46), we have

$$1 = \sum_{i=1}^{m} p_i = \sum_{i=1}^{m} \frac{w_i}{1 + \lambda(2\hat{G}(W_i) - 1)} \tag{3.47}$$

and

$$1 = 2 \sum_{i=1}^{m} \hat{G}(W_i)p_i = \sum_{i=1}^{m} \frac{2\hat{G}(W_i)w_i}{1 + \lambda(2\hat{G}(W_i) - 1)} \tag{3.48}$$

which give

$$0 = \sum_{i=1}^{m} \frac{2\hat{G}(W_i)w_i}{1 + \lambda(2\hat{G}(W_i) - 1)} - \sum_{i=1}^{m} \frac{w_i}{1 + \lambda(2\hat{G}(W_i) - 1)} = \sum_{i=1}^{m} \frac{w_i(2\hat{G}(W_i) - 1)}{1 + \lambda(2\hat{G}(W_i) - 1)}. \tag{3.49}$$

Thus, from (3.46) and (3.49), a candidate for the solution of (3.41) is given by

$$\hat{p}_i^0 = \frac{w_i}{1 + \lambda_0(2\hat{G}(W_i) - 1)}, \qquad 1 \leq i \leq m, \tag{3.50}$$

with λ_0 as a solution of equation (3.49). The following lemma shows that this candidate (3.50) is the unique solution for (3.41).

Lemma 3.4. *For optimization problem (3.41), the probability of the following events tend to one as $n \to \infty$.*

(i) Equation (3.49) has a unique solution on interval

$$\left(\frac{-1}{2\hat{G}(W_{(m)}) - 1}, \frac{-1}{2\hat{G}(W_{(1)}) - 1} \right), \tag{3.51}$$

where

$$W_{(1)} = \min\{W_1, \ldots, W_m\} \qquad and \qquad W_{(m)} = \max\{W_1, \ldots, W_m\}; \tag{3.52}$$

(ii) $\hat{\mathbf{p}}^0 = (\hat{p}_1^0, \ldots, \hat{p}_m^0)$ in (3.50) is the unique solution of (3.41).

Proof of (i)   First, we show that

$$\hat{G}(W_{(1)}) < \frac{1}{2} < \hat{G}(W_{(m)}), \qquad \text{in probability.} \tag{3.53}$$

As reviewed in Section 1.3, under certain regularity conditions we have $\|\hat{G} - F_X\| \overset{a.s.}{\to} 0$ for right censored data (1.16) [40], doubly censored data (1.17) [6] and [17], interval censored Case 1 or Case 2 data (1.18)$-$(1.19) [15], and partly interval censored data (1.21)$-$(1.22) [18], which implies that

$$\|\hat{G} - F_X\| \overset{P}{\longrightarrow} 0, \quad \text{as } n \to \infty \tag{3.54}$$

for the aforementioned types of censored data. Also, from (2.2), we have $\hat{G}(W_1^X -) = 0$ and $\hat{G}(W_{m_1}^X) = 1$ and from (3.42), we have $W_{(m)} \geq W_{m_1}^X$; in turn, we have

$$\hat{G}(W_{(m)}) \geq \hat{G}(W_{m_1}^X) = 1 > 1/2,$$

because $\hat{G}$ is a non-decreasing function. We choose $\delta > 0$ such that

$$0 < F_X(m_X - 2\delta) \leq F_X(m_X - \delta) < F_X(m_X) = 1/2, \tag{3.55}$$

where m_X is the median of X, that is $F_X(m_X) = 1/2$. Since $\hat{G}(W_1^X - \delta) = 0$, we have

$$P\{W_1^X \geq m_X - \delta\} = P\{W_1^X - \delta \geq m_X - 2\delta\} \leq P\{\hat{G}(W_1^X - \delta) \geq \hat{G}(m_X - 2\delta)\}$$

$$= P\{0 \geq \hat{G}(m_X - 2\delta)\} \to 0, \quad \text{as } n \to \infty, \tag{3.56}$$

because, from (3.54),

$$\hat{G}(m_X - 2\delta) \overset{P}{\longrightarrow} F_X(m_X - 2\delta) > 0, \quad \text{as } n \to \infty.$$

Also, from (3.55), we choose $\epsilon > 0$ such that $F_X(m_X - \delta) + \epsilon < 1/2$. Then, we have that for $\Delta = \hat{G}(m_X - \delta) - F_X(m_X - \delta)$,

$$
\begin{aligned}
P\left\{W_1^X < m_X - \delta\right\} &\leq P\left\{\hat{G}(W_1^X) \leq \hat{G}(m_X - \delta)\right\} \\
&= P\left\{\hat{G}(W_1^X) \leq F_X(m_X - \delta) + \Delta\right\} \\
&= P\left\{\hat{G}(W_1^X) \leq F_X(m_X - \delta) + \Delta \,\big|\, |\Delta| < \epsilon\right\} \\
&\quad + P\left\{\hat{G}(W_1^X) \leq F_X(m_X - \delta) + \Delta \,\big|\, |\Delta| \geq \epsilon\right\} \\
&\leq P\left\{\hat{G}(W_1^X) \leq F_X(m_X - \delta) + \epsilon\right\} + P\left\{|\Delta| \geq \epsilon\right\} \\
&\leq P\left\{\hat{G}(W_1^X) < 1/2\right\} + P\left\{|\Delta| \geq \epsilon\right\}.
\end{aligned}
\tag{3.57}
$$

Since $\lim_{n\to\infty} P\{W_1^X < m_X - \delta\} = 1$, from (3.54) and (3.56)$-$(3.57), we have

$$
\lim_{n\to\infty} P\{\hat{G}(W_1^X) < 1/2\} = 1.
\tag{3.58}
$$

Thus, we have $\hat{G}(W_1^X) < 1/2$ in probability, and from (3.42), we have $W_{(1)} \leq W_1^X$. In turn, in probability, $\hat{G}(W_{(1)}) \leq \hat{G}(W_1^X) < 1/2$ because $\hat{G}$ is a non-decreasing function. Hence, we have (3.53).

Note that from (3.53), we have $2\hat{G}(W_{(1)}) - 1 < 0$ and $2\hat{G}(W_{(m)}) - 1 > 0$ in probability. For $2\hat{G}(W_i) - 1 > 0$, we have:

$$
\begin{aligned}
2\hat{G}(W_{(1)}) - 1 \leq 2\hat{G}(W_i) - 1 \quad &\Leftrightarrow \quad -(2\hat{G}(W_{(1)}) - 1) \geq -(2\hat{G}(W_i) - 1) \\
&\Leftrightarrow \quad -\frac{1}{2\hat{G}(W_i) - 1} \geq -\frac{1}{2\hat{G}(W_{(1)}) - 1},
\end{aligned}
$$

66

and

$$2\hat{G}(W_{(m)}) - 1 \geq 2\hat{G}(W_i) - 1 \quad \Leftrightarrow \quad -(2\hat{G}(W_{(m)}) - 1) \leq -(2\hat{G}(W_i) - 1)$$

$$\Leftrightarrow \quad -\frac{1}{2\hat{G}(W_i) - 1} \leq -\frac{1}{2\hat{G}(W_{(m)}) - 1}.$$

Then, from (3.46), $w_i > 0$, and requirement $p_i > 0$ for all $1 \leq i \leq m$, we require for all $1 \leq i \leq m$,

$$1 + \lambda(2\hat{G}(W_i) - 1) > 0 \quad \Leftrightarrow \quad \lambda(2\hat{G}(W_i) - 1) > -1$$

$$\Leftrightarrow \quad \max_{\substack{1 \leq i \leq m \\ 2\hat{G}(W_i)-1>0}} \frac{-1}{2\hat{G}(W_i) - 1} < \lambda < \min_{\substack{1 \leq i \leq m \\ 2\hat{G}(W_i)-1<0}} \frac{-1}{2\hat{G}(W_i) - 1}$$

$$\Leftrightarrow \quad \frac{-1}{2\hat{G}(W_{(m)}) - 1} < \lambda < \frac{-1}{2\hat{G}(W_{(1)}) - 1}.$$

Thus, to have all $p_i > 0$ in (3.46), we are only interested in a solution of (3.49) on interval (3.51). From (3.49), we denote

$$g(\lambda) \;=\; \sum_{i=1}^{m} \frac{w_i(2\hat{G}(W_i) - 1)}{1 + \lambda(2\hat{G}(W_i) - 1)} \;=\; 0. \tag{3.59}$$

Then, since $w_i > 0$ for $1 \leq i \leq m$, and (3.53), we have that for any λ in interval (3.51),

$$g'(\lambda) = -\sum_{i=1}^{m} \frac{w_i(2\hat{G}(W_i) - 1)^2}{\left[1 + \lambda(2\hat{G}(W_i) - 1)\right]^2} < 0. \tag{3.60}$$

which implies that $g(\lambda)$ is strictly decreasing on (3.51). Letting $\lambda_1 = \frac{-1}{2\hat{G}(W_{(m)})-1}$ and $\lambda_2 = \frac{-1}{2\hat{G}(W_{(1)})-1}$, we have

$$\lim_{\lambda \to \lambda_1^+} g(\lambda) = \lim_{\lambda \to \lambda_1^+} \sum_{i=1}^{m} \frac{w_i(2\hat{G}(W_i) - 1)}{1 + \lambda(2\hat{G}(W_i) - 1)} = \lim_{\lambda \to \lambda_1^+} \sum_{2\hat{G}(W_i)\neq 1} \frac{w_i}{\lambda + \frac{1}{2\hat{G}(W_i)-1}}$$

$$= \sum_{\substack{2\hat{G}(W_i)\neq 1 \\ W_i=W_{(m)}}} \lim_{\lambda \to \lambda_1^+} \frac{w_i}{\lambda + \frac{1}{2\hat{G}(W_{(m)})-1}} + \sum_{\substack{2\hat{G}(W_i)\neq 1 \\ W_i\neq W_{(m)}}} \lim_{\lambda \to \lambda_1^+} \frac{w_i}{\lambda + \frac{1}{2\hat{G}(W_i)-1}} \tag{3.61}$$

$$= \infty + C_1 = \infty,$$

and

$$\lim_{\lambda \to \lambda_2^-} g(\lambda) = \lim_{\lambda \to \lambda_2^-} \sum_{i=1}^{m} \frac{w_i(2\hat{G}(W_i) - 1)}{1 + \lambda(2\hat{G}(W_i) - 1)} = \lim_{\lambda \to \lambda_2^-} \sum_{2\hat{G}(W_i) \neq 1} \frac{w_i}{\lambda + \frac{1}{2\hat{G}(W_i)-1}}$$

$$= \sum_{\substack{2\hat{G}(W_i) \neq 1 \\ W_i = W_{(1)}}} \lim_{\lambda \to \lambda_2^-} \frac{w_i}{\lambda + \frac{1}{2\hat{G}(W_{(1)})-1}} + \sum_{\substack{2\hat{G}(W_i) \neq 1 \\ W_i \neq W_{(1)}}} \lim_{\lambda \to \lambda_2^-} \frac{w_i}{\lambda + \frac{1}{2\hat{G}(W_i)-1}} \tag{3.62}$$

$$= -\infty + C_2 = -\infty,$$

where C_1 and C_2 are finite constants. Thus, (3.49) has a unique solution on (3.51).

Proof of (ii) We prove that $\hat{\mathbf{p}}^0 = (\hat{p}_1^0, \ldots, \hat{p}_m^0)$, given by (3.50), is the unique solution to (3.41) by verifying the KKT conditions in Theorem 4.3.8 of Bazarra, Sherali, and Shetty (1993; page 164) as follows. From (3.41) and (3.43), we introduce the notations

$$\mathcal{F} = \left\{ \mathbf{p} \ \middle| \ 0 < p_i < 1, \ 1 \leq i \leq m; \ \sum_{i=1}^{m} p_i = 1; \ 2\sum_{i=1}^{m} \hat{G}(W_i)p_i = 1 \right\}, \tag{3.63}$$

and

$$h(\mathbf{p}) = n \sum_{i=1}^{m} w_i \log p_i. \tag{3.64}$$

The Hessian matrix (Bazarra, Sherali, and Shetty, 1993; page 90) of $h(\mathbf{p})$ given in (3.64), exists on set $\mathcal{F}$ and is given by

$$\frac{\partial^2 h(\mathbf{p})}{\partial p_i \partial p_j} = \begin{cases} -\dfrac{nw_i}{p_i^2} & \text{if } i = j \\[2mm] 0 & \text{if } i \neq j. \end{cases} \implies \mathbf{H}_h = \text{diag}\left\{ -\frac{nw_1}{p_1^2}, \ldots, -\frac{nw_m}{p_m^2} \right\}. \tag{3.65}$$

Since $\mathbf{H}_h$ is a diagonal matrix with diagonal elements $-\frac{nw_i}{p_i^2} < 0$, $1 \leq i \leq m$, for $\mathbf{p} \in \mathcal{F}$, $\mathbf{H}_h$ is negative definite on $\mathcal{F}$. Note that $\mathcal{F}$ is a convex set because for any $\mathbf{p}$, $\mathbf{q} \in \mathcal{F}$ and

68

$\mathbf{r} = \lambda\mathbf{p} + (1-\lambda)\mathbf{q}$ with any $\lambda \in (0,1)$ we have

$$0 < r_i = \lambda p_i + (1-\lambda)q_i < \lambda + (1-\lambda) = 1, \text{ for all } 1 \le i \le m;$$

$$\sum_{i=1}^{m} r_i = \lambda \sum_{i=1}^{m} p_i + (1-\lambda) \sum_{i=1}^{m} q_i = \lambda + (1-\lambda) = 1; \tag{3.66}$$

$$2\sum_{i=1}^{m} \hat{G}(W_i)r_i = \lambda\left(2\sum_{i=1}^{m} \hat{G}(W_i)p_i\right) + (1-\lambda)\left(2\sum_{i=1}^{m} \hat{G}(W_i)q_i\right) = \lambda + (1-\lambda) = 1.$$

Thus, function $h(\mathbf{p})$ is strictly concave on $\mathcal{F}$ by Theorem 3.3.8 of Bazarra, Sherali, and Shetty (1993; pages 93 and 79). To verify the conditions in Theorem 4.3.8 of Bazarra, Sherali, and Shetty (1993; page 164), note that $X_p = \{\mathbf{p} \mid 0 < p_i < 1, \ 1 \le i \le m\}$ is a nonempty open set in $\mathbb{R}^{>}$, and that $h(\mathbf{p})$, $h_1(\mathbf{p}) = 1 - \sum_{i=1}^{m} p_i$, and $h_2(\mathbf{p}) = 1 - 2\sum_{i=1}^{m} \hat{G}(W_i)p_i$ are each from $\mathbb{R}^{>} \to \mathbb{R}$. Since $\hat{\mathbf{p}}^0 \in X_p$ satisfies constraints $h_1(\hat{\mathbf{p}}^0) = 0$ and $h_2(\hat{\mathbf{p}}^0) = 0$, $\hat{\mathbf{p}}^0$ is a feasible solution for (3.41) (Bazarra, Sherali, and Shetty, 1993; page 99). Also, note that with $v_1 = n(1-\lambda_0)$ and $v_2 = n\lambda_0$, the KKT conditions are satisfied because

$$\boldsymbol{\nabla}h(\hat{\mathbf{p}}^0) + v_1\boldsymbol{\nabla}h_1(\hat{\mathbf{p}}^0) + v_2\boldsymbol{\nabla}h_2(\hat{\mathbf{p}}^0)$$

$$= \begin{pmatrix} nw_1/\hat{p}_1^0 \\ \vdots \\ nw_m/\hat{p}_m^0 \end{pmatrix} + n(1-\lambda_0)\begin{pmatrix} -1 \\ \vdots \\ -1 \end{pmatrix} + n\lambda_0\begin{pmatrix} -2\hat{G}(W_1) \\ \vdots \\ -2\hat{G}(W_m) \end{pmatrix}$$

$$= \begin{pmatrix} n + n\lambda_0(2\hat{G}(W_1) - 1) \\ \vdots \\ n + n\lambda_0(2\hat{G}(W_m) - 1) \end{pmatrix} + \begin{pmatrix} -n + n\lambda_0 - 2n\lambda_0\hat{G}(W_1) \\ \vdots \\ -n + n\lambda_0 - 2n\lambda_0\hat{G}(W_m) \end{pmatrix} = \mathbf{0}.$$

Since $h(\mathbf{p})$ is concave and differentiable on $\mathcal{F}$, $h(\mathbf{p})$ is pseudoconcave on $\mathcal{F}$ (Bazarra, Sherali, and Shetty, 1993; page 116). Note that both h_1 and h_2 are linear functions which means that both h_1 and h_2 are quasiconvex and quasiconcave on $\mathcal{F}$ (Bazarra, Sherali, and Shetty,

1993; pages 116 and 118). Thus, by Theorems 3.4.2 and 4.3.8 of Bazarra, Sherali, and Shetty (1993; pages 101 and 164), $\hat{\mathbf{p}}^0$ is the unique global optimal solution to (3.41).

Hence, by Lemma 3.4, the numerator of (3.36) is given by

$$L(\gamma_{00}, \hat{F}_n^0) = \gamma_{00}{}^{n_0} \prod_{i=1}^{m} \left(\hat{p}_i^0\right)^{nw_i}, \tag{3.67}$$

where for $\hat{p}_i^0$'s given by (3.50), we have $\hat{F}_n^0(x) = \sum_{i=1}^{m} \hat{p}_i^0 I\{W_i \leq x\}$.

From (3.40) and (3.67), weighted empirical likelihood ratio test statistic (3.36) for test (3.35) is given by

$$R_0 = R(\mathbf{O}_n) = \frac{L(\gamma_{00}, \hat{F}_n^0)}{L(\hat{\gamma}, \hat{F}_n)} = \frac{\gamma_{00}{}^{n_0} \prod_{i=1}^{m} \left(\hat{p}_i^0\right)^{nw_i}}{\hat{\gamma}^{n_0} \prod_{i=1}^{m} w_i^{nw_i}} = \left(\frac{\gamma_{00}}{\hat{\gamma}}\right)^{n_0} \prod_{i=1}^{m} \left(\frac{\hat{p}_i^0}{w_i}\right)^{nw_i}, \tag{3.68}$$

and we reject H_0 in (3.35) when $R(\mathbf{O}_n)$ is small. In order to determine the rejection region, we need to know the asymptotic distribution of $-2\log R_0$ under H_0, which is to be studied in the future and is expected to be a scaled χ^2 distribution.

3.3.2 Confidence Intervals

To obtain the weighted empirical likelihood ratio based confidence interval (WELRBCI) for γ_0, we first notice that $R(\mathbf{O}_n)$ in (3.36) can be written in the form of (1.34) by the following way. Since the last constraint in (3.34) reflects the relation between γ and F under Two-Sample Accelerated Life Model (1.5), γ and F are linked through the statistical functional $T(\cdot)$ of F given by

$$T(F) \equiv [\text{ solution of } \varphi_n(\gamma; F) \doteq 0] \iff \gamma = T(F), \tag{3.69}$$

70

where F is given by (2.7) and we have

$$\varphi_n(\gamma; F) \equiv \varphi_n(\gamma; \mathbf{p}) = 2 \sum_{i=1}^{m} \hat{G}(W_i^{\gamma}) p_i - 1, \tag{3.70}$$

because $\varphi_n(\gamma; F) = 0$ is equivalent to the last constraint equation in (3.34). Thus, for F given by (2.7), $R(\mathbf{O}_n)$ in (3.36) is given by

$$R(\mathbf{O}_n) = \sup_{(\gamma, F) \in H_0} \frac{L(\gamma, F)}{L(\hat{\gamma}, \hat{F}_n)} \overset{(3.69)}{=} \sup_{\substack{(\gamma, F) \in H_0 \\ \gamma = T(F)}} \frac{L(T(F), F)}{L(\hat{\gamma}, \hat{F}_n)}$$

$$\overset{(3.35)}{=} \sup_{T(F) = \gamma_{00}} \frac{L(T(F), F)}{L(\hat{\gamma}, \hat{F}_n)} = \sup_{T(F) = \gamma_{00}} \frac{\mathcal{L}(F)}{\mathcal{L}(\hat{F}_n)}, \tag{3.71}$$

where

$$\mathcal{L}(F) \equiv L(T(F), F), \tag{3.72}$$

and from (3.39) and (3.69), we have

$$\mathcal{L}(\hat{F}_n) = L(T(\hat{F}_n), \hat{F}_n) = L(\hat{\gamma}, \hat{F}_n). \tag{3.73}$$

From (3.71), the acceptance region for (3.35), analogously to (1.38), is given by

$$\{\mathbf{O}_n \mid R(\mathbf{O}_n) \geq c\} = \left\{ \mathbf{O}_n \,\middle|\, \sup_{T(F) = \gamma_{00}} \frac{\mathcal{L}(F)}{\mathcal{L}(\hat{F}_n)} \geq c \right\} = \left\{ \mathbf{O}_n \,\middle|\, \sup_{T(F) = \gamma_{00}} \lambda(F) \geq c \right\}, \tag{3.74}$$

for some predetermined $0 < c < 1$, where similar to (1.39), from (2.6), (3.40) and (3.71)$-$(3.73), we have

$$\lambda(F) = \frac{\mathcal{L}(F)}{\mathcal{L}(\hat{F}_n)} = \left(\frac{T(F)}{\hat{\gamma}} \right)^{n_0} \prod_{i=1}^{m} \left(\frac{p_i}{w_i} \right)^{n w_i}. \tag{3.75}$$

Analogous to (1.40), from (2.7) and (3.74)$-$(3.75), the weighted empirical likelihood ratio confidence region for $\gamma_0 = \gamma_{00}$ is given by

$$S = \left\{ \gamma \,\middle|\, \sup_{T(F) = \gamma} \lambda(F) \geq c \right\} = \{ T(\mathbf{p}) \mid \mathbf{p} \in E_c \}, \tag{3.76}$$

provided that $T(\mathbf{p}) = T(F)$ is continuous, where

$$E_c = \left\{ \mathbf{p} \ \middle| \ 0 < p_i < 1, 1 \le i \le m; \ \sum_{i=1}^{m} p_i = 1; \ \left(\frac{T(\mathbf{p})}{\hat{\gamma}}\right)^{n_0} \prod_{i=1}^{m} \left(\frac{p_i}{w_i}\right)^{nw_i} \ge c \right\}. \tag{3.77}$$

3.3.2.1 Simplification of $\varphi_n(\gamma; \mathbf{p})$:

Note that the continuity assumption on $T(\cdot)$ is required for the last equality in (3.76) to hold. To study if $T(\mathbf{p})$ in (3.69) is continuous, we note that the solution to $\varphi_n(\gamma; \mathbf{p}) \doteqdot 0$ exists by the definition of the notation "$\doteqdot$" introduced in (3.14), and we simplify $\varphi_n(\gamma; \mathbf{p})$ in (3.70) as follows:

$$\begin{aligned}
\varphi_n(\gamma; \mathbf{p}) &= 2\sum_{i=1}^{m_1} \hat{G}(W_i^X) p_i + 2\sum_{i=1}^{m_0} \hat{G}(\gamma W_i^Y) p_{i+m_1} - 1 \tag{3.78} \\
&= 2\sum_{i=1}^{m_1} \hat{G}(W_i^X) p_i - 1 + 2\sum_{i=1}^{m_0} p_{i+m_1} \left[\sum_{j=1}^{m_1} \hat{p}_j^X I\{W_j^X \le \gamma W_i^Y\}\right] \\
&= 2\sum_{i=1}^{m_1} \hat{G}(W_i^X) p_i - 1 + 2\sum_{i=1}^{m_0} \sum_{j=1}^{m_1} p_{i+m_1} \hat{p}_j^X I\{(W_j^X/W_i^Y) \le \gamma\} \\
&= h(\mathbf{p}) + \sum_{i=1}^{N} \Delta_i(\mathbf{p}) I\{U_i \le \gamma\}, \tag{3.79}
\end{aligned}$$

where U_i's are as in (3.30)$-$(3.31), $\Delta_i(\mathbf{p})$ denotes the size of the jump at U_i and we have

$$h(\mathbf{p}) = 2\sum_{i=1}^{m_1} \hat{G}(W_i^X) p_i - 1. \tag{3.80}$$

Note that (3.79) implies that for any fixed $\mathbf{p}$, $\varphi_n(\gamma; \mathbf{p})$ is a monotone non-decreasing piecewise step-function, thus the solution to $\varphi_n(\gamma; \mathbf{p}) \doteqdot 0$ is not unique. This means that $T(\mathbf{p})$ as written in (3.69) is not well-defined.

72

In order to write $T(\mathbf{p})$ as a well-defined function, we study the behavior of $\varphi_n(\gamma; \mathbf{p})$ for fixed $\mathbf{p}$ as follows. For fixed $\mathbf{p}$, from $\hat{G}(0) = 0$ and $\hat{G}(\infty) = 1$, respectively, we have in (3.78)

$$\varphi_n(0; \mathbf{p}) = 2\sum_{i=1}^{m_1} \hat{G}(W_i^X)p_i - 1 = h(\mathbf{p}) \tag{3.81}$$

$$\varphi_n(\infty; \mathbf{p}) = 2\sum_{i=1}^{m_1} \hat{G}(W_i^X)p_i + 2\sum_{i=1}^{m_0} p_{i+m_1} - 1 = h(\mathbf{p}) + \sum_{i=1}^{N} \Delta_i(\mathbf{p}). \tag{3.82}$$

Since $\varphi_n(\gamma; \mathbf{p})$ is a monotone non-decreasing function in γ, we know that for fixed $\mathbf{p}$ there are three possible scenarios for $\varphi_n(\gamma; \mathbf{p})$ at the end points of interval $(0, \infty)$ given as follows:

$$E_1 = \{\mathbf{p} \mid \varphi_n(0; \mathbf{p}) > 0 \text{ and } \varphi_n(\infty; \mathbf{p}) > 0\}; \tag{3.83}$$

$$E_2 = \{\mathbf{p} \mid \varphi_n(0; \mathbf{p}) < 0 \text{ and } \varphi_n(\infty; \mathbf{p}) < 0\}; \tag{3.84}$$

$$E_3 = \{\mathbf{p} \mid \varphi_n(0; \mathbf{p}) < 0 \text{ and } \varphi_n(\infty; \mathbf{p}) > 0\}. \tag{3.85}$$

Note that from (3.79), for any fixed $\mathbf{p}$, $\varphi_n(\gamma; \mathbf{p})$ only has jumps at U_i's. Thus, for $\mathbf{p} \in E_1$, as shown in Figure 3.4, all line segments lie above the γ-axis and $|\varphi_n(U_1; \mathbf{p})| < |\varphi_n(U_i; \mathbf{p})|$ for all $2 \leq i \leq N$. Similarly, for $\mathbf{p} \in E_2$, as shown in Figure 3.4, all line segments lie below the γ-axis and $|\varphi_n(U_N; \mathbf{p})| < |\varphi_n(U_i; \mathbf{p})|$ for all $1 \leq i \leq N - 1$. For $\mathbf{p} \in E_3$, there are three scenarios as shown in Figures 3.5$-$3.7, where the use of γ', γ_L, and γ_U are as in Figures 3.1$-$3.3. Thus, we define $T(\mathbf{p})$ in (3.69) as follows: Denoting

$$\gamma_L = \inf\left\{\gamma \mid \varphi_n(\gamma; \mathbf{p}) = \varphi_n(\gamma'-; \mathbf{p})\right\}, \quad \gamma_U = \sup\left\{\gamma \mid \varphi_n(\gamma; \mathbf{p}) = \varphi_n(\gamma'; \mathbf{p})\right\}, \tag{3.86}$$

we have

$$T(\mathbf{p}) \equiv \begin{cases} U_1 & \text{if } \mathbf{p} \in E_1 \\[2mm] U_N & \text{if } \mathbf{p} \in E_2 \text{ or } \mathbf{p} \in E_3 \text{ with } \gamma' = U_N \\[2mm] \tilde{\gamma} & \text{if } \mathbf{p} \in E_3 \text{ with } \gamma' < U_N, \end{cases} \qquad (3.87)$$

where

$$\tilde{\gamma} = \begin{cases} \dfrac{\gamma_L + \gamma'}{2} & \text{if } |\varphi_n(\gamma_L; \mathbf{p})| < |\varphi_n(\gamma'; \mathbf{p})| \\[3mm] \dfrac{\gamma' + \gamma_U}{2} & \text{if } |\varphi_n(\gamma_L; \mathbf{p})| > |\varphi_n(\gamma'; \mathbf{p})| \\[3mm] \dfrac{\gamma_L + \gamma_U}{2} & \text{if } |\varphi_n(\gamma_L; \mathbf{p})| = |\varphi_n(\gamma'; \mathbf{p})|. \end{cases} \qquad (3.88)$$
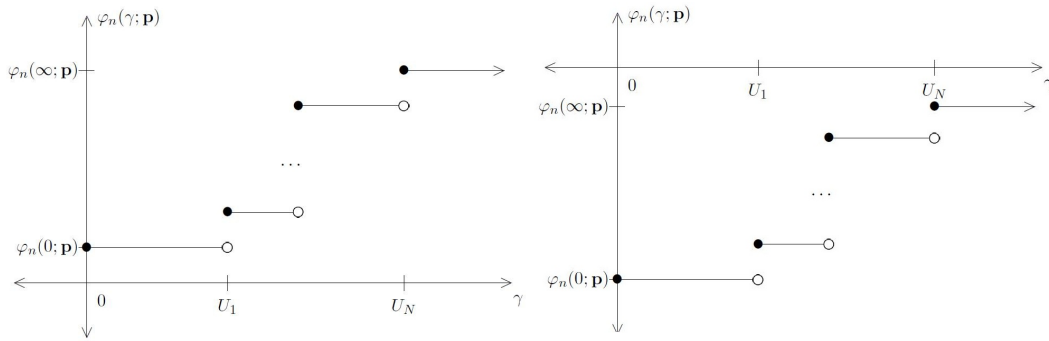
Figure 3.4: Scenarios for $\varphi(\gamma; \mathbf{p})$ for fixed $p \in E_1 \cup E_2$
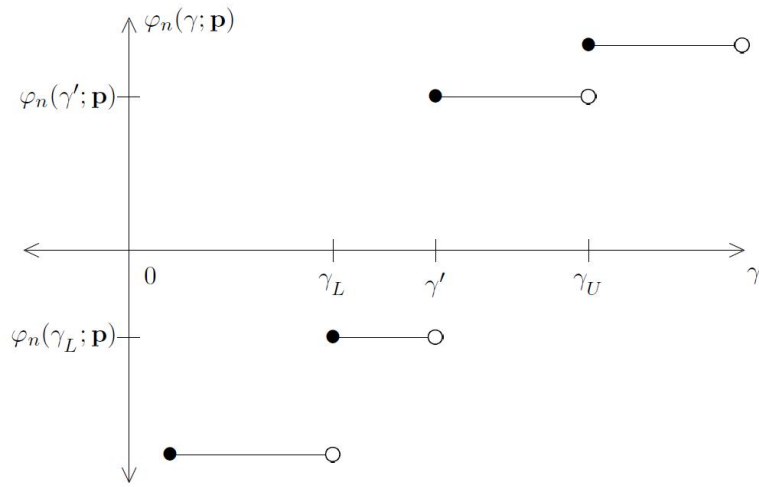
Figure 3.5: Scenarios for $\varphi(\gamma; \mathbf{p})$ for fixed $p \in E_3$: $\varphi(U_1; p) < 0$ $\varphi(U_{N-1}; p) > 0$
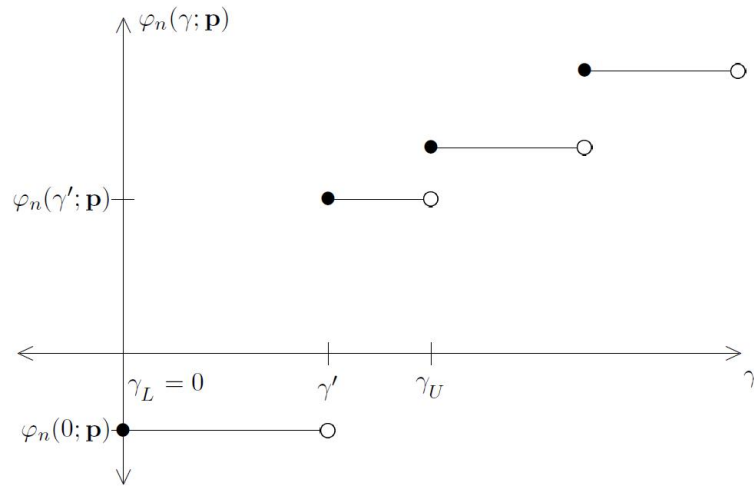
Figure 3.6: Scenarios for $\varphi(\gamma; \mathbf{p})$ for fixed $p \in E_3$: $\varphi(U_1; p) > 0$
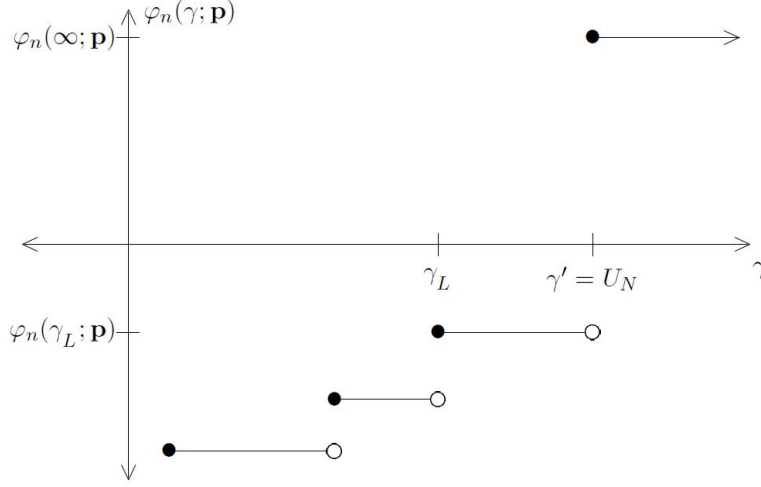
Figure 3.7: Scenarios for $\varphi(\gamma; \mathbf{p})$ for fixed $p \in E_3$: $\varphi(U_{N-1}; p) < 0$

3.3.2.2 Continuity of $T(\cdot)$:

To show that $T(\mathbf{p})$ is continuous, letting $\mathbf{p}^{(k)} \to \mathbf{p}^{(0)}$ as $k \to \infty$ we need to show that

$$T(\mathbf{p}^{(k)}) \; \to \; T(\mathbf{p}^{(0)}), \qquad \text{as } k \to \infty. \tag{3.89}$$

As follows, we establish (3.89) for $\mathbf{p}^{(0)} \in E_i$, for $i = 1, 2, 3$, respectively.

For $\mathbf{p}^{(0)} \in E_1$, by (3.87) we have $T(\mathbf{p}^{(0)}) = U_1$. From (3.79) and (3.81), for any $\mathbf{p}$ and any $0 < \gamma_1 < U_1$, we have

$$\varphi_n(\gamma_1; \mathbf{p}) \; = \; h(\mathbf{p}) + \sum_{i=1}^{N} \Delta_i(\mathbf{p}) I\{U_i \le \gamma_1\} \; = \; h(\mathbf{p}) \; = \; \varphi_n(0; \mathbf{p}). \tag{3.90}$$

From (3.80), we know that $h(\mathbf{p})$ is a continuous function in $\mathbf{p}$, thus by (3.79) and (3.83) we have

$$\lim_{k \to \infty} \varphi_n(\gamma_1; \mathbf{p}^{(k)}) \; = \; \lim_{k \to \infty} h(\mathbf{p}^{(k)}) \; = \; h(\mathbf{p}^{(0)}) \; = \; \varphi_n(0; \mathbf{p}^{(0)}) > 0, \tag{3.91}$$

which implies that there exists K_1 such that $\varphi_n(\gamma_1; \mathbf{p}^{(k)}) = \varphi_n(0; \mathbf{p}^{(k)}) > 0$ for $k \geq K_1$.

Hence, by (3.83), we have $\mathbf{p}^{(k)} \in E_1$ for all $k \geq K_1$ which gives

$$\lim_{k \to \infty} T(\mathbf{p}^{(k)}) = U_1 = T(\mathbf{p}^{(0)}). \tag{3.92}$$

For $\mathbf{p}^{(0)} \in E_2$, by (3.87) we have $T(\mathbf{p}^{(0)}) = U_N$. From (3.79) and (3.82), for any $\mathbf{p}$ and any $U_N < \gamma_2 < \infty$, we have

$$\varphi_n(\gamma_2; \mathbf{p}) = h(\mathbf{p}) + \sum_{i=1}^{N} \Delta_i(\mathbf{p}) I\{U_i \leq \gamma_2\} = h(\mathbf{p}) + \sum_{i=1}^{N} \Delta_i(\mathbf{p}) = \varphi_n(\infty; \mathbf{p}). \tag{3.93}$$

From (3.78), we know that $\varphi_n(\infty; \mathbf{p})$ is a continuous function in $\mathbf{p}$, thus by (3.84) we have

$$\lim_{k \to \infty} \varphi_n(\gamma_2; \mathbf{p}^{(k)}) = \lim_{k \to \infty} \varphi_n(\infty; \mathbf{p}^{(k)}) = \varphi_n(\infty; \mathbf{p}^{(0)}) < 0, \tag{3.94}$$

which implies that there exists K_2 such that $\varphi_n(\gamma_2; \mathbf{p}^{(k)}) = \varphi_n(\infty; \mathbf{p}^{(k)}) < 0$ for $k \geq K_2$.

Thus, by (3.84), we have $\mathbf{p}^{(k)} \in E_2$ for all $k \geq K_2$ which gives

$$\lim_{k \to \infty} T(\mathbf{p}^{(k)}) = U_N = T(\mathbf{p}^{(0)}). \tag{3.95}$$

For $\mathbf{p}^{(0)} \in E_3$, from Figure 3.3 we need to consider two cases: (i) $\gamma_0' = U_N$; (ii) $\gamma_0' < U_N$, where γ_0' is γ' in Figure 3.3 which corresponds to $\mathbf{p}^{(0)}$.

Case (i): For $\gamma' = U_N$, by (3.87) we have $T(\mathbf{p}^{(0)}) = U_N$, and from Figure 3.3 (c) we know that $\gamma_L = U_{N-1}$ with

$$\varphi_n(U_{N-1}; \mathbf{p}^{(0)}) < 0 \qquad \text{and} \qquad \varphi_n(U_N; \mathbf{p}^{(0)}) > 0. \tag{3.96}$$

From (3.78), we know that for any fixed γ, $\varphi_n(\gamma; \mathbf{p})$ is continuous in $\mathbf{p}$. Thus, (3.96) implies

$$\lim_{k \to \infty} \varphi_n(U_{N-1}; \mathbf{p}^{(k)}) = \varphi_n(U_{N-1}; \mathbf{p}^{(0)}) < 0$$

$$\lim_{k \to \infty} \varphi_n(U_N; \mathbf{p}^{(k)}) = \varphi_n(U_N; \mathbf{p}^{(0)}) > 0. \tag{3.97}$$

From (3.97), there exists K such that for any $k \geq K$, we have $\varphi_n(U_{N-1}; \mathbf{p}^{(k)}) < 0$ and $\varphi_n(U_N; \mathbf{p}^{(k)}) > 0$ for all $k \geq K$. Thus, we have $T(\mathbf{p}^{(k)}) \equiv U_N$ for all $k \geq K$, which implies $\lim_{k \to \infty} T(\mathbf{p}^{(k)}) = T(\mathbf{p}^{(0)}) = U_N$.

Case (ii): For $\gamma' < U_N$, by (3.87) we have $T(\mathbf{p}^{(0)}) = \tilde{\gamma}_0$ where $\tilde{\gamma}_0$ corresponds to $\gamma_L^{(0)}$ and γ_0' for $\mathbf{p}^{(0)}$ in Figure 3.3 (a)(b). Thus, we have

$$\varphi_n(\gamma_L^{(0)}; \mathbf{p}^{(0)}) < 0 \qquad \text{and} \qquad \varphi_n(\gamma_0'; \mathbf{p}^{(0)}) > 0 \tag{3.98}$$

which implies

$$\begin{aligned} \lim_{k \to \infty} \varphi_n(\gamma_L^{(0)}; \mathbf{p}^{(k)}) &= \varphi_n(\gamma_L^{(0)}; \mathbf{p}^{(0)}) < 0 \\ \lim_{k \to \infty} \varphi_n(\gamma_0'; \mathbf{p}^{(k)}) &= \varphi_n(\gamma_0'; \mathbf{p}^{(0)}) > 0. \end{aligned} \tag{3.99}$$

The proof follows from the arguments line-by-line after (3.97) for Case (i).

The following two lemmas establish properties of S in (3.76) and establish a relationship between S and R_0 in (3.71).

Lemma 3.5. S is an interval that satisfies $S = [W_L, W_U]$ where

$$W_L = \min_{\mathbf{p} \in E_c} T(\mathbf{p}) \qquad and \qquad W_U = \max_{\mathbf{p} \in E_c} T(\mathbf{p}). \tag{3.100}$$

Proof First, we let $y \in S$ in (3.76), which implies that $y = T(\mathbf{p})$ for some $\mathbf{p}' \in E_c$ in (3.77), that is, from (3.69)$-$(3.70) y is a solution of $2 \sum_{i=1}^{m} \hat{G}(W_i^y) p_i' = 1$ for some $\mathbf{p}' \in E_c$. Then, we have

$$W_L = \min_{\mathbf{p} \in E_C} T(\mathbf{p}) \leq y \leq \max_{\mathbf{p} \in E_C} T(\mathbf{p}) = W_U \qquad \Rightarrow \qquad y \in [W_L, W_U].$$

78

Now, we let $y \in [W_L, W_U]$. Since function $T(\mathbf{p})$ is continuous on E_c, $T(\mathbf{p})$ attains its minimum and maximum on E_c. Hence, we have $W_L = T(\mathbf{p}^L)$ and $W_U = T(\mathbf{p}^U)$ for some $\mathbf{p}^L$, $\mathbf{p}^U \in E_c$, which gives

$$T(\mathbf{p}^L) = W_L \leq y \leq W_U = T(\mathbf{p}^U).$$

Now, it suffices to show that $y = T(\mathbf{p}')$ for some $\mathbf{p}' \in E_c$. Consider

$$h(\lambda) = T\big((1 - \lambda)\mathbf{p}^L + \lambda \mathbf{p}^U\big), \quad 0 \leq \lambda \leq 1.$$

Then, we have $h(0) = T(\mathbf{p}^L) \leq y \leq T(\mathbf{p}^U) = h(1)$. Note that $h(\lambda)$ is a continuous function since T is continuous. Then, by the Intermediate Value Theorem, we have that there exists a $\lambda' \in [0, 1]$ such that $h(0) \leq y = h(\lambda') \leq h(1)$, which implies that

$$T(\mathbf{p}^L) \leq y = T(\mathbf{p}') = \sum_{i=1}^{m} p_i' \hat{W}_i \leq T(\mathbf{p}^U), \text{ where } \mathbf{p}' = (1 - \lambda')\mathbf{p}^L + \lambda' \mathbf{p}^U. \tag{3.101}$$

We complete the proof by showing $\mathbf{p}' \in E_c$.

Since $0 \leq p_i^L \leq 1$ and $0 \leq p_i^U \leq 1$, $1 \leq i \leq m$, we have

$$0 \leq p_i' = (1 - \lambda')p_i^L + \lambda' p_i^U \leq (1 - \lambda^*) + \lambda' = 1, \ i = 1, \ldots, m$$

and since $\sum_{i=1}^{m} p_i^L = 1$ and $\sum_{i=1}^{m} p_i^U = 1$, we have

$$\sum_{i=1}^{m} p_i' = \sum_{i=1}^{m} \Big[(1 - \lambda')p_i^L + \lambda' p_i^U\Big] = (1 - \lambda')\sum_{i=1}^{m} p_i^L + \lambda' \sum_{i=1}^{m} p_i^U = (1 - \lambda') + \lambda' = 1.$$

It only remains to show that $\left(\frac{T(\mathbf{p}')}{\hat{\gamma}}\right)^{n_0} \prod_{i=1}^{m} \left(\frac{p_i'}{w_i}\right)^{nw_i} \geq c$. To show this we need to show that $D(\mathbf{p}) = \left(\frac{T(\mathbf{p})}{\hat{\gamma}}\right)^{n_0} \prod_{i=1}^{m} \left(\frac{p_i}{w_i}\right)^{nw_i}$ is a convex function. For this we need to check that $\nabla_p^2(D) \geq 0$.

Lemma 3.6. *We have $W_L \leq \gamma_{00} \leq W_U$ if and only if $R_0 \geq c$.*

79

Proof   Suppose $W_L \leq \gamma_{00} \leq W_U$. Then, we have that $\gamma_{00} = T(F)$, that is, γ_{00} is a solution

to $2 \sum_{i=1}^{m} \hat{G}(W_i) p_i' = 1$ for some $\mathbf{p}' \in E_c$. Since $\mathbf{p}' \in E_c$, we have

$$\left(\frac{\gamma_{00}}{\hat{\gamma}} \right)^{n_0} \prod_{i=1}^{m} \left(\frac{p_i'}{w_i} \right)^{nw_i} \geq c, \ \sum_{i=1}^{m} p_i' = 1, \ 0 \leq p_i' \leq 1, \ \text{for } i = 1, \ldots, m.$$

Then, for $\mathcal{F}$ in (3.63), from (3.68) we have

$$c \leq \left(\frac{\gamma_{00}}{\hat{\gamma}} \right)^{n_0} \prod_{i=1}^{m} \left(\frac{p_i'}{w_i} \right)^{nw_i} \leq \sup_{\mathbf{p} \in F_c} \left(\frac{\gamma_{00}}{\hat{\gamma}} \right)^{n_0} \prod_{i=1}^{m} \left(\frac{p_i}{w_i} \right)^{nw_i} = R_0.$$

Now suppose $R_0 \geq c$. To show that $W_L \leq \gamma_{00} \leq W_U$, we first show that

$\max_{\mathbf{p} \in \mathcal{F}} \left(\frac{\gamma_{00}}{\hat{\gamma}} \right)^{n_0} \prod_{i=1}^{m} \left(\frac{p_i}{w_i} \right)^{nw_i}$ is attained on the set $\mathcal{F}$. To do so, we denote the closure of $\mathcal{F}$

as:

$$\mathcal{F}' = \left\{ \mathbf{p} \ \middle| \ 0 \leq p_i \leq 1, \ 1 \leq i \leq m; \ \sum_{i=1}^{m} p_i = 1; \ 2 \sum_{i=1}^{m} \hat{G}(W_i) p_i = 1 \right\}, \qquad (3.102)$$

Note that $\mathcal{F}'$ is a subset of $\mathbb{R}^{>}$ and is bounded because of the constraint $0 \leq p_i \leq 1$,

$i = 1, \ldots, m$. Also note that if $\mathbf{p}^{(k)} \to \mathbf{p}^{(0)}$, as $k \to \infty$ for a sequence $\mathbf{p}^{(k)} \in F_c$, we have

$$2 \sum_{i=1}^{m} \hat{G}(W_i) p_i^{(k)} = 1, \ \sum_{i=1}^{m} p_i^{(k)} = 1, \ 0 \leq p_i^{(k)} \leq 1, \ \text{for } i = 1, \ldots, m, \qquad (3.103)$$

and we have

$$1 = \lim_{k \to \infty} 2 \sum_{i=1}^{m} \hat{G}(W_i) p_i^{(k)} = 2 \sum_{i=1}^{m} \hat{G}(W_i) \lim_{k \to \infty} p_i^{(k)} = 2 \sum_{i=1}^{m} \hat{G}(W_i) p_i^{(0)}$$

$$1 = \lim_{k \to \infty} \sum_{i=1}^{m} p_i^{(k)} = \sum_{i=1}^{m} \lim_{k \to \infty} p_i^{(k)} = \sum_{i=1}^{m} p_i^{(0)}$$

$$0 \leq \lim_{k \to \infty} p_i^{(k)} = p_i^{(0)} \leq 1, i = 1, \ldots, m,$$

which implies that $\mathbf{p}^{(0)} \in \mathcal{F}'$. Thus, $\mathcal{F}'$ is closed; in turn, we know that $\mathcal{F}'$ is compact.

Since function $f(\mathbf{p}) = \left(\frac{\gamma_{00}}{\hat{\gamma}} \right)^{n_0} \prod_{i=1}^{m} \left(\frac{p_i}{w_i} \right)^{nw_i}$ is continuous on $\mathcal{F}'$, $f(\mathbf{p})$ attains its maximum

80

for some $\mathbf{p}'$ on $\mathcal{F}'$. Recall by the argument in $(3.33)-(3.34)$, we cannot have $p_i = 0$ or $p_i = 1$ for any $i = 1, \ldots, m$. Hence, we must have that $f(\mathbf{p})$ attains its maximum for some $\mathbf{p}'$ on $\mathcal{F}$ and we have:

$$c \leq R_0 = \max_{\mathbf{p} \in \mathcal{F}} \prod_{i=1}^{m} \left(\frac{\gamma_{00}}{\hat{\gamma}} \right)^{n_0} \left(\frac{p_i}{w_i} \right)^{nw_i} = \left(\frac{\gamma_{00}}{\hat{\gamma}} \right)^{n_0} \prod_{i=1}^{m} \left(\frac{p_i'}{w_i} \right)^{nw_i}.$$

Since $\mathbf{p}' \in \mathcal{F}$, we have

$$\left(\frac{\gamma_{00}}{\hat{\gamma}} \right)^{n_0} \prod_{i=1}^{m} \left(\frac{p_i'}{w_i} \right)^{nw_i} \geq c, \ 2\sum_{i=1}^{m} \hat{G}(W_i) p_i' = 1, \ \sum_{i=1}^{m} p_i' = 1, \ 0 \leq p_i' \leq 1, \ \text{for } i = 1, \ldots, m.$$

Hence, γ_{00} is a solution of $2 \sum_{i=1}^{m} \hat{G}(W_i) p_i' = 1$ with $\mathbf{p}' \in E_c$; in turn, $\gamma_{00} \in \{T(F) \mid \mathbf{p} \in E_c\}$, which implies that $\gamma_{00} \in [W_L, W_U]$.

From Lemma 3.6, we have

$$P\{W_L \leq \gamma_{00} \leq W_U\} = P\{-2 \log R_0 \leq -2 \log c\}. \tag{3.104}$$

In turn, the constant c in (3.76) is determined by the limiting distribution of $-2 \log R_0$, which, as mentioned in Section 3.3.1, is to be studied in the future and is expected to be a scaled χ^2 distribution.

3.3.2.3 Computation of $[W_L, W_U]$:

From (3.100), W_L and W_U can be obtained by solving the following optimization problems, respectively:

$$
\begin{cases}
\text{Minimize / Maximize} \quad & T(\mathbf{p}) \\[2ex]
\text{subject to:} \quad & 0 < p_i < 1, \ 1 \leq i \leq m; \ \sum_{i=1}^{m} p_i = 1; \\[2ex]
& \left(\frac{T(\mathbf{p})}{\hat{\gamma}} \right)^{n_0} \prod_{i=1}^{m} \left(\frac{p_i}{w_i} \right)^{n w_i} \geq c.
\end{cases}
\tag{3.105}
$$

The solution to the above optimization problem is to be studied in the future.

CHAPTER 4
ESTIMATION AND TESTS ON TREATMENT MEAN

In this chapter, we construct hypothesis tests and estimators for the mean μ_X of the treatment distribution F_X in Two-Sample Accelerated Life Model (1.5) based on estimator $\hat{\gamma}_E$ in (3.2), estimator $\hat{\gamma}_R$ in (3.28) and the Weighted Empirical Likelihood Method, respectively. The methods developed in this chapter are applicable in a unified way to those different types of censored data described in Section 1.3. The organization of this chapter is as follows. Section 4.1 gives point estimators $\hat{\mu}_E$ and $\hat{\mu}_R$ for μ_X based on $\hat{\gamma}_E$ and $\hat{\gamma}_R$, respectively. Section 4.2 constructs tests and confidence intervals for μ_X based on estimators $\hat{\mu}_E$ and $\hat{\mu}_R$, respectively. Section 4.3 constructs Weighted Empirical Likelihood Ratio based tests and confidence intervals for μ_X. Comparison of the estimators in this chapter by simulation studies is given in Chapter 5.

4.1   Point Estimators

Since $\hat{G}$ given in (2.2) is an estimator for F_X, a natural point estimator $\hat{\mu}_X$ for μ_X is given by

$$\hat{\mu}_X = \int x \, d\hat{G}(x) = \sum_{i=1}^{m_1} \hat{p}_i^X W_i^X. \tag{4.1}$$

But such an estimator is only based on the first sample in (2.1), thus it is less efficient. Since $\hat{F}_n(\,\cdot\,; \hat{\eta})$ in (2.21) is an estimator for F_X that is calculated using both samples in (2.1), a

more efficient point estimator for μ_X is naturally given by

$$
\begin{aligned}
\hat{\mu}(\hat{\eta}) &\equiv \int x \, d\hat{F}_n(x; \hat{\eta}) = \int x \, d\left(\rho_1 \hat{G}(x) + \rho_0 \hat{H}(x/\hat{\eta})\right) \\
&= \rho_1 \int x \, d\hat{G}(x) + \rho_0 \int x \, d\hat{H}(x/\hat{\eta}) \\
&= \rho_1 \hat{\mu}_X + \rho_0 \hat{\eta} \int x/\hat{\eta} \, d\hat{H}(x/\hat{\eta}) = \rho_1 \hat{\mu}_X + \rho_0 \hat{\eta} \hat{\mu}_Y,
\end{aligned}
\tag{4.2}
$$

where the notation $\hat{\mu}(\hat{\eta})$ indicates the dependance on a consistent estimator $\hat{\eta}$ for γ_0. Recall from Chapter 3 that estimators $\hat{\gamma}_E$ and $\hat{\gamma}_R$ given by (3.2) and (3.28), respectively, are consistent estimators for γ_0. Hence, we have two point estimators for μ_X by plugging $\hat{\gamma}_E$ and $\hat{\gamma}_R$, respectively, into (4.2) as follows:

$$
\hat{\mu}_E \equiv \hat{\mu}_X = \hat{\mu}(\hat{\gamma}_E) = \rho_1 \hat{\mu}_X + \rho_0 \left(\hat{\mu}_X/\hat{\mu}_Y\right) \hat{\mu}_Y = (\rho_1 + \rho_0)\hat{\mu}_X,
\tag{4.3}
$$

$$
\hat{\mu}_R \equiv \hat{\mu}(\hat{\gamma}_R) = \rho_0 \hat{\mu}_X + \hat{\gamma}_R \rho_1 \hat{\mu}_Y.
\tag{4.4}
$$

Also note that $\hat{\mu}_R$ in (4.4) is a more efficient estimator than $\hat{\mu}_E$ since it uses both samples in (2.1). The following theorem establishes some asymptotic properties of point estimator $\hat{\mu}_R$ under assumptions (AS5)−(AS6) in Theorem 3.1, which, by Remark 3.1, hold for all types of censored data considered in this dissertation.

Theorem 4.1. *Assume that (AS5)−(AS6) from Theorem 3.1 hold. Then,*

$$
\sqrt{n} \left(\hat{\mu}_R - \mu_X\right) \xrightarrow{D} N(0, \sigma_{\mu_R}^2), \quad \text{as } n \to \infty.
\tag{4.5}
$$

Proof From (4.4), we have:

$$\sqrt{n}(\hat{\mu}_R - \mu_X) = \sqrt{n}\left(\rho_1\hat{\mu}_X + \rho_0\hat{\gamma}_R\hat{\mu}_Y - \mu_X\right)$$

$$= \sqrt{n}\left(\rho_1\hat{\mu}_X + \rho_0\hat{\gamma}_R\hat{\mu}_Y - \rho_0\hat{\gamma}_R\mu_Y + \rho_0\hat{\gamma}_R\mu_Y - \rho_0\mu_X - \rho_1\mu_X\right)$$

$$= \rho_1\sqrt{n}\left(\hat{\mu}_X - \mu_X\right) + \rho_0\hat{\gamma}_R\sqrt{n}\left(\hat{\mu}_Y - \mu_Y\right) + \rho_0\sqrt{n}\left(\hat{\gamma}_R\mu_Y - \mu_X\right)$$

$$\xrightarrow{D} N(0, \sigma_{\mu_R}^2),$$

by (AS5)−(AS6).

4.2   Normal-Approximation Based Tests and Confidence Intervals

In this section, we construct tests and confidence intervals for μ_X based on a point estimator $\hat{\mu}$ for μ_X which satisfies:

$$\sqrt{n}\left(\hat{\mu} - \mu_X\right) \xrightarrow{D} N(0, \sigma^2), \quad \text{as } n \to \infty. \tag{4.6}$$

Note that from (AS5) in Theorem 3.1 and (4.5), we know that both estimators $\hat{\mu}_E$ and $\hat{\mu}_R$ in (4.3)−(4.4) satisfy the assumption in (4.6). Thus, all procedures in this section apply to both of these estimators. In turn, the procedures are also applicable to the various types of censored data considered in this dissertation.

4.2.1 Hypothesis Tests

We consider the following hypothesis test on the mean of the treatment group:

$$H_0 : \mu_X = \mu_0 \quad \text{vs.} \quad H_1 : \ \mu_X \neq \mu_0. \tag{4.7}$$

Based on point estimator $\hat{\mu}$, in practice we reject H_0 in (4.7) if $|\hat{\mu} - \mu_0| > c$ for some predetermined $c > 0$. For level of significance $0 < \alpha < 1$, we may determine c in practice via (4.6) as follows:

$$
\begin{aligned}
\alpha &= P\{\text{Type I Error}\} = P\{\text{reject } H_0 \mid H_0 \text{ is true}\} \\
&= P\left\{|\hat{\mu} - \mu_0| > c \mid \mu_X = \mu_0\right\} = P\left\{\left|\frac{\hat{\mu} - \mu_X}{\sigma/\sqrt{n}}\right| > \frac{c}{\sigma/\sqrt{n}} \;\middle|\; \mu_X = \mu_0\right\} \\
&\approx P\left\{|Z| > \frac{c}{\sigma/\sqrt{n}}\right\} = 2P\left\{Z > \frac{c}{\sigma/\sqrt{n}}\right\},
\end{aligned} \tag{4.8}
$$

which gives

$$\frac{c}{\sigma/\sqrt{n}} = z_{\alpha/2} \quad \Rightarrow \quad c = \frac{\sigma}{\sqrt{n}} z_{\alpha/2}. \tag{4.9}$$

To estimate σ in (4.9) in practice for point estimators $\hat{\mu}_E$ and $\hat{\mu}_R$, we may use the following bootstrap procedures [12], respectively, which are valid for all of the types of censored data considered in this dissertation because of (4.6).

Estimation of σ_{μ_E}:

Step 1. Generate bootstrap sample $\mathbf{O}_1^{X*}, \ldots, \mathbf{O}_{n_1}^{X*}$ as in the first sample in (2.26).

Step 2. Compute $\hat{G}^*$ based on (2.2) using the first bootstrap sample in (2.26).

Step 3. Compute $\hat{\mu}_E^* = \int x \, d\hat{G}^*(x)$.

Step 4. Repeat Steps $1 - 3$ B times to obtain $\hat{\mu}_E^*(b)$, $b = 1, \ldots B$, where B is usually

chosen to be 1000. The bootstrap estimate for standard error σ_{μ_E} is given by

$$\hat{se}_{\hat{\mu}_E} = \left(\sum_{b=1}^{B} \frac{\left(\hat{\mu}_E^*(b) - \frac{1}{B} \sum_{i=1}^{B} \hat{\mu}_E^*(i) \right)^2}{B-1} \right)^{1/2}. \tag{4.10}$$

Estimation of σ_{μ_R}:

Step 1. Generate bootstrap samples $\mathbf{O}_1^{X*}, \ldots, \mathbf{O}_{n_1}^{X*}$ and $\mathbf{O}_1^{Y*}, \ldots, \mathbf{O}_{n_0}^{Y*}$ as in (2.26).

Step 2. Compute $\hat{G}^*$ and $\hat{H}^*$ as in (2.2) using the bootstrap samples (2.26).

Step 3. Compute $\hat{\mu}_X^* = \int x \, d\hat{G}^*(x)$ and $\hat{\mu}_Y^* = \int x \, d\hat{H}^*(x)$ as in (3.3).

Step 4. Compute $\hat{\gamma}_R^*$, which is the solution of $g(\gamma; \hat{G}^*, \hat{H}^*) \doteq 0$.

Step 5. Compute $\hat{\mu}_R^*$ as in (4.4).

Step 6. Repeat Steps $1 - 5$ B times to obtain $\hat{\mu}_R^*(b)$, $b = 1, \ldots B$, where B is usually

chosen to be 1000. The bootstrap estimate for standard error σ_{μ_R} in Theorem 4.1

is given by

$$\hat{se}_{\hat{\mu}_R} = \left(\sum_{b=1}^{B} \frac{\left(\hat{\mu}_R^*(b) - \frac{1}{B} \sum_{i=1}^{B} \hat{\mu}_R^*(i) \right)^2}{B-1} \right)^{1/2}. \tag{4.11}$$

4.2.2 Confidence Intervals

A $(1-\alpha)100\%$ confidence interval for μ_X based on point estimator $\hat{\mu}$ is constructed as follows. From (4.6), we have

$$
\begin{aligned}
1 - \alpha &= P\left\{|Z| \le z_{\alpha/2}\right\} \approx P\left\{-z_{\alpha/2} \le \frac{\hat{\mu} - \mu_X}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right\} \\
&= P\left\{\hat{\mu} - \frac{\sigma}{\sqrt{n}}z_{\alpha/2} \le \mu_X \le \hat{\mu} + \frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right\}.
\end{aligned}
$$

Thus, an approximated $(1 - \alpha)100\%$ confidence interval for μ_X based on $\hat{\mu}$ is given by

$$
\hat{\mu} \pm \frac{\hat{\sigma}}{\sqrt{n}}z_{\alpha/2}, \tag{4.12}
$$

where $\hat{\sigma}$ is an estimator for σ, e.g., possible estimators for σ_{μ_E} and σ_{μ_R} are given in the above bootstrap procedures $(4.10)-(4.11)$, respectively. Some simulation results on this are presented in Chapter 5.

4.3 Weighted Empirical Likelihood Ratio Tests and Confidence Intervals

Recall that the weighted empirical likelihood function $L(\gamma, F)$ for (γ_0, F_X) is given by (2.6). Since $\hat{\gamma} = \hat{\gamma}_R$ in (3.39) is the WELMLE for γ_0 and is a consistent estimator for γ_0, which, by Remark 3.2, is applicable for all types of censored data considered in this dissertation, we consider the following likelihood function:

$$
L(\hat{\gamma}, F) = \hat{\gamma}^{n_0} \prod_{i=1}^{m} p_i^{nw_i}, \tag{4.13}
$$

where, as in (2.5) and (2.7) we have:

$$
\begin{cases}
(\hat{W}_1, \ldots, \hat{W}_m) \;=\; (W_1^{\hat{\gamma}}, \ldots, W_m^{\hat{\gamma}}) \\[2mm]
F(x) = \sum_{i=1}^m p_i I\{\hat{W}_i \le x\}, \;\; \text{with } p_i = F(\hat{W}_i) - F(\hat{W}_i-), \;\; 1 \le i \le m.
\end{cases}
\tag{4.14}
$$

In the next two subsections, via likelihood function $L(\hat{\gamma}, F)$ in (4.13) we construct Weighted Empirical Likelihood Ratio based tests and confidence intervals for μ_X with some simulation results presented in Chapter 5.

4.3.1 Hypothesis Tests

Under Two-Sample Accelerated Life Model (1.5), consider hypothesis test (4.7). For likelihood function $L(\hat{\gamma}, F)$ in (4.13), the weighted empirical likelihood ratio function is given by

$$
R(F; \hat{\gamma}) = \frac{L(\hat{\gamma}, F)}{L(\hat{\gamma}, \hat{F}_n)} = \frac{\hat{\gamma}^{n_0} \prod_{i=1}^m p_i^{nw_i}}{\hat{\gamma}^{n_0} \prod_{i=1}^m w_i^{nw_i}} = \prod_{i=1}^m \left(\frac{p_i}{w_i} \right)^{nw_i},
\tag{4.15}
$$

where $\hat{F}_n$ is given in (3.39) and w_i's are given in (2.5). Thus, the weighted empirical likelihood ratio test statistic for (4.7) analogous to (1.34) is given by:

$$
R_0 \;=\; R(\mathbf{O}_n) \;=\; \sup_{F \in H_0} R(F; \hat{\gamma}) \;=\; \sup_{T(F)=\mu_0} \prod_{i=1}^m \left(\frac{p_i}{w_i} \right)^{nw_i}
\tag{4.16}
$$

where $\mathbf{O}_n$ is given by (3.37), F is given by (4.14) and

$$
T(F) = \int x \, dF(x) = \sum_{i=1}^m p_i \hat{W}_i.
\tag{4.17}
$$

To obtain an expression for R_0 in (4.16), note that from (4.13)$-$(4.14) and (4.16)$-$(4.17)

we need to solve the following optimization problem

$$\begin{cases} \text{Maximize} & \prod_{i=1}^{m} \left(\frac{p_i}{w_i} \right)^{nw_i} \\ \text{subject to: } 0 \le p_i \le 1, \ 1 \le i \le m; \ \sum_{i=1}^{m} p_i = 1; \ \sum_{i=1}^{m} p_i \hat{W}_i = \mu_0. \end{cases} \quad (4.18)$$

By the same argument used in (2.8)$-$(2.9), we know that for F given in (4.14), optimization

problem (4.18) is equivalent to:

$$\begin{cases} \text{Maximize} & \prod_{i=1}^{m} \left(\frac{p_i}{w_i} \right)^{nw_i} \\ \text{subject to: } 0 < p_i < 1, \ 1 \le i \le m; \ \sum_{i=1}^{m} p_i = 1; \ \sum_{i=1}^{m} p_i \hat{W}_i = \mu_0. \end{cases} \quad (4.19)$$

The following lemma gives the solution to optimization problem (4.19).

Lemma 4.2. *Under Two-Sample Accelerated Life Model (1.5), assume*

(AS7) $\quad \|\hat{F}_n(\cdot; \hat{\gamma}) - F_X\| \xrightarrow{P} 0, \ as \ n \to \infty.$

Then, the solution to (4.19) is given by:

$$\tilde{p}_i = \frac{w_i}{1 + \lambda_0(\hat{W}_i - \mu_0)}, \qquad 1 \le i \le m, \quad (4.20)$$

where λ_0 is the unique solution to equation

$$\sum_{i=1}^{m} \frac{w_i(\hat{W}_i - \mu_0)}{1 + \lambda(\hat{W}_i - \mu_0)} = 0 \quad (4.21)$$

on the interval

$$J = \left(\frac{-1}{\hat{W}_{(m)} - \mu_0}, \frac{-1}{\hat{W}_{(1)} - \mu_0} \right), \quad (4.22)$$

where

$$\hat{W}_{(1)} = \min\{\hat{W}_1, \dots, \hat{W}_m\} \qquad and \qquad \hat{W}_{(m)} = \max\{\hat{W}_1, \dots, \hat{W}_m\}. \quad (4.23)$$

Remark 4.1. *We discuss (AS7) in Lemma 4.2 as follows.* As reviewed in Section 1.3, under certain regularity conditions we have $||\hat{G} - F_X|| \overset{a.s.}{\to} 0$, as $n_1 \to \infty$ and $||\hat{H} - F_Y|| \overset{a.s.}{\to} 0$, as $n_0 \to \infty$ for right censored data (1.16) [40], doubly censored data (1.17) [6] and [17], interval censored Case 1 or Case 2 data (1.18)$-$(1.19) [15], and partly interval censored data (1.21)$-$(1.22) [18], which implies that

$$||\hat{G} - F_X|| \overset{P}{\longrightarrow} 0, \quad \text{as } n_1 \to \infty \qquad \text{and} \qquad ||\hat{H} - F_Y|| \overset{P}{\longrightarrow} 0, \quad \text{as } n_0 \to \infty \qquad (4.24)$$

for the aforementioned types of censored data. Then, under Two-Sample Accelerated Life Model (1.5), from (2.21), (4.24), and the fact that $\hat{\gamma}$ is a consistent estimator for γ_0, we have:

$$||\hat{F}_n(\cdot;\hat{\gamma}) - F_X|| = \sup_x |\rho_1 \hat{G}(x) + \rho_0 \hat{H}(x/\hat{\gamma}) - F_X(x)|$$

$$= \sup_x |\rho_1 \hat{G}(x) - \rho_1 F_X(x) + \rho_0 \hat{H}(x/\hat{\gamma}) - \rho_0 F_X(x)|$$

$$\leq \rho_1 \sup_x |\hat{G}(x) - F_X(x)| + \rho_0 \sup_x |\hat{H}(x/\hat{\gamma}) - F_X(x)|$$

$$= \rho_1 ||\hat{G} - F_X|| + \rho_0 \sup_x |\hat{H}(x/\hat{\gamma}) - F_Y(x/\gamma_0)|$$

$$= \rho_1 ||\hat{G} - F_X|| + \rho_0 \sup_x |\hat{H}(x/\hat{\gamma}) - F_Y(x/\gamma_0)|$$

$$\leq \rho_1 ||\hat{G} - F_X|| + \rho_0 \sup_x |\hat{H}(x/\hat{\gamma}) - F_Y(x/\hat{\gamma})| + \rho_0 \sup_x |F_Y(x/\hat{\gamma}) - F_Y(x/\gamma_0)|$$

$$\leq \rho_1 ||\hat{G} - F_X|| + \rho_0 ||\hat{H} - F_Y|| + \rho_0 \sup_x |F_Y(x/\hat{\gamma}) - F_Y(x/\gamma_0)|$$

$$\overset{P}{\to} 0, \quad \text{as } n \to \infty,$$

provided that d.f. F_Y is uniformly continuous.

Proof To solve optimization problem (4.19), we note that for all $1 \leq i \leq m$, we have

$$\log \prod_{i=1}^{m} \left(\frac{p_i}{w_i}\right)^{nw_i} = n \sum_{i=1}^{m} w_i \log p_i - n \sum_{i=1}^{m} w_i \log w_i, \qquad (4.25)$$

Thus, to find a candidate for the solution using the Lagrange Multipliers, we denote

$$\mathscr{H}(\mathbf{p}, \beta, \lambda) = \sum_{=} \log{}_i + \beta \left[- \sum_{=} {}_i \right] + \lambda \left[\mu - \sum_{=} {}_i \hat{\mathscr{W}} \right], \tag{4.26}$$

then, we have for $1 \le i \le m$,

$$0 = \frac{\partial \mathscr{H}}{\partial p_i} = \frac{n w_i}{p_i} - \beta - n \lambda \hat{W}_i \quad \Rightarrow \quad p_i = \frac{n w_i}{\beta + n \lambda \hat{W}_i}. \tag{4.27}$$

From (2.3), (2.5) and the last two constraints in (4.19), we have

$$\beta p_i = n w_i - n \lambda p_i \hat{W}_i \quad \Rightarrow \quad \beta \sum_{i=1}^{m} p_i = n \lambda \sum_{i=1}^{m} p_i \hat{W}_i \tag{4.28}$$

$$\Rightarrow \quad \beta = n - n \lambda \mu_0 \quad \Rightarrow \quad \beta = n(1 - \lambda \mu_0).$$

Plugging β in (4.28) into (4.27), we have

$$p_i = \frac{n w_i}{n - n \lambda \mu_0 + n \lambda \hat{W}_i} = \frac{w_i}{1 + \lambda(\hat{W}_i - \mu_0)}, \quad 1 \le i \le m. \tag{4.29}$$

From the constraints in (4.19) and p_i in (4.29), we have

$$1 = \sum_{i=1}^{m} p_i = \sum_{i=1}^{m} \frac{w_i}{1 + \lambda(\hat{W}_i - \mu_0)} \tag{4.30}$$

and

$$\mu_0 = \sum_{i=1}^{m} p_i \hat{W}_i = \sum_{i=1}^{m} \frac{w_i \hat{W}_i}{1 + \lambda(\hat{W}_i - \mu_0)} \tag{4.31}$$

which give

$$0 = \sum_{i=1}^{m} \frac{w_i \hat{W}_i}{1 + \lambda(\hat{W}_i - \mu_0)} - \mu_0 \sum_{i=1}^{m} \frac{w_i}{1 + \lambda(\hat{W}_i - \mu_0)} = \sum_{i=1}^{m} \frac{w_i(\hat{W}_i - \mu_0)}{1 + \lambda(\hat{W}_i - \mu_0)} \equiv g(\lambda). \tag{4.32}$$

Thus, from (4.29) and (4.32) a candidate for the solution of (4.19) is given by

$$p_i = \frac{n w_i}{n - n \lambda_0 \mu_0 + n \lambda_0 \hat{W}_i} = \frac{w_i}{1 + \lambda_0(\hat{W}_i - \mu_0)}. \tag{4.33}$$

with λ_0 as a solution of equation (4.32).

Next, we show that equation (4.80) has a unique solution $\lambda = \lambda_0$ in the interval

$$J = \left(\frac{-1}{\hat{W}_{(m)} - \mu_0}, \frac{-1}{\hat{W}_{(1)} - \mu_0} \right). \tag{4.34}$$

First, we note that since $\hat{F}_n$ is non-degenerate, (AS7) implies that as $n \to \infty$,

$$\hat{W}_{(1)} < \mu_0 < \hat{W}_{(m)}, \qquad \text{in probability.} \tag{4.35}$$

which implies that, in probability, we have $\hat{W}_{(1)} - \mu_0 < 0$ and $\hat{W}_{(m)} - \mu_0 > 0$ and we have

$$\hat{W}_{(1)} - \mu_0 \leq \hat{W}_i - \mu_0 \quad \Leftrightarrow \quad -(\hat{W}_{(1)} - \mu_0) \geq -(\hat{W}_i - \mu_0) \quad \Leftrightarrow \quad -\frac{1}{\hat{W}_i - \mu_0} \geq \frac{-1}{\hat{W}_{(1)} - \mu_0},$$

and for $\hat{W}_i - \mu_0 > 0$, we have

$$\hat{W}_{(m)} - \mu_0 \geq \hat{W}_i - \mu_0 \quad \Leftrightarrow \quad -(\hat{W}_{(m)} - \mu_0) \leq -(\hat{W}_i - \mu_0) \quad \Leftrightarrow \quad -\frac{1}{\hat{W}_i - \mu_0} \leq \frac{-1}{\hat{W}_{(m)} - \mu_0}.$$

Then, from (4.29), $w_i > 0$, and requirement $p_i > 0$ for all $1 \leq i \leq m$, we require for all $1 \leq i \leq m$,

$$1 + \lambda(\hat{W}_i - \mu_0) > 0 \quad \Leftrightarrow \quad \lambda(\hat{W}_i - \mu_0) > -1$$

$$\Leftrightarrow \quad \max_{\substack{1 \leq i \leq m \\ \hat{W}_i - \mu_0 > 0}} \frac{-1}{\hat{W}_i - \mu_0} < \lambda < \min_{\substack{1 \leq i \leq m \\ \hat{W}_i - \mu_0 < 0}} \frac{-1}{\hat{W}_i - \mu_0}$$

$$\Leftrightarrow \quad \frac{-1}{\hat{W}_{(m)} - \mu_0} < \lambda < \frac{-1}{\hat{W}_{(1)} - \mu_0}.$$

Thus, to have all $p_i > 0$ in (4.29), we are only interested in a solution of (4.19) on interval J. From (4.21), we denote

$$g(\lambda) = \sum_{i=1}^{m} \frac{w_i(\hat{W}_i - \mu_0)}{1 + \lambda(\hat{W}_i - \mu_0)} = 0 \tag{4.36}$$

93

Now, from (4.19), $w_i > 0$ for $1 \leq i \leq m$, and (4.35), we have that for any $\lambda \in J$,

$$g'(\lambda) = -\sum_{i=1}^{m} \frac{w_i(\hat{W}_i - \mu_0)^2}{\left[1 + \lambda(\hat{W}_i - \mu_0)\right]^2} < 0, \tag{4.37}$$

which implies that $g(\lambda)$ is strictly decreasing on J. Letting $\lambda_1 = \frac{-1}{\hat{W}_{(m)} - \mu_0}$ and $\lambda_2 = \frac{-1}{\hat{W}_{(1)} - \mu_0}$,

we have

$$
\begin{aligned}
\lim_{\lambda \to \lambda_1^+} g(\lambda) &= \lim_{\lambda \to \lambda_1^+} \sum_{i=1}^{m} \frac{w_i(\hat{W}_i - \mu_0)}{1 + \lambda(\hat{W}_i - \mu_0)} = \lim_{\lambda \to \lambda_1^+} \sum_{\hat{W}_i \neq \mu_0} \frac{w_i}{\lambda + \frac{1}{\hat{W}_i - \mu_0}} \\
&= \sum_{\hat{W}_i \neq \mu_0, \hat{W}_i = \hat{W}_{(m)}} \lim_{\lambda \to \lambda_1^+} \frac{w_i}{\lambda + \frac{1}{\hat{W}_{(m)} - \mu_0}} + \sum_{\hat{W}_i \neq \mu_0, \hat{W}_i \neq \hat{W}_{(m)}} \lim_{\lambda \to \lambda_1^+} \frac{w_i}{\lambda + \frac{1}{\hat{W}_i - \mu_0}} \\
&= \infty + C_2 = \infty,
\end{aligned}
$$

and

$$
\begin{aligned}
\lim_{\lambda \to \lambda_2^-} g(\lambda) &= \lim_{\lambda \to \lambda_2^-} \sum_{i=1}^{m} \frac{w_i(\hat{W}_i - \mu_0)}{1 + \lambda(\hat{W}_i - \mu_0)} = \lim_{\lambda \to \lambda_2^-} \sum_{\hat{W}_i \neq \mu_0} \frac{w_i}{\lambda + \frac{1}{\hat{W}_i - \mu_0}} \\
&= \sum_{\hat{W}_i \neq \mu_0, \hat{W}_i = \hat{W}_{(1)}} \lim_{\lambda \to \lambda_2^-} \frac{w_i}{\lambda + \frac{1}{\hat{W}_{(1)} - \mu_0}} + \sum_{\hat{W}_i \neq \mu_0, \hat{W}_i \neq \hat{W}_{(1)}} \lim_{\lambda \to \lambda_2^-} \frac{w_i}{\lambda + \frac{1}{\hat{W}_i - \mu_0}} \\
&= -\infty + C_3 = -\infty,
\end{aligned}
$$

where C_2 and C_3 are finite constants. Thus, (4.80) has a unique solution on J.

We prove that $\hat{\mathbf{p}} = (\hat{p}_1, \ldots, \hat{p}_m)$, given by

$$\hat{p}_i = \frac{w_i}{1 + \lambda_0(\hat{W}_i - \mu_0)}, \quad 1 \leq i \leq m, \tag{4.38}$$

is the unique solution to (4.19) by verifying the KKT conditions in Theorem 4.3.8 of Bazarra,

Sherali, and Shetty (page 164) as follows. Let

$$h(\mathbf{p}) = \gamma^{n_0} \prod_{i=1}^{m} p_i^{nw_i}, \tag{4.39}$$

and let

$$\tilde{F}_c = \left\{ \mathbf{p} \ \middle| \ 0 < p_i < 1, \ 1 \le i \le m; \ \sum_{i=1}^{m} p_i = 1; \ 2\sum_{i=1}^{m} \hat{G}(W_i)p_i = 1 \right\}. \qquad (4.40)$$

The Hessian matrix (Bazarra, Sherali, and Shetty, 1993; page 90) of $h(\mathbf{p})$, given in (4.39), exists on set F_c and is given by

$$\frac{\partial^2 h(\mathbf{p})}{\partial p_i \partial p_j} = \begin{cases} -\dfrac{nw_i}{p_i^2} & \text{if } i = j \\[2ex] 0 & \text{if } i \ne j. \end{cases} \qquad \Longrightarrow \qquad \mathbf{H}_h = \text{diag}\left\{ -\frac{nw_1}{p_1^2}, \dots, -\frac{nw_m}{p_m^2} \right\}. \qquad (4.41)$$

Since $\mathbf{H}_h$ is a diagonal matrix with diagonal elements $-\frac{nw_i}{p_i^2} < 0$, $1 \le i \le m$, for $\mathbf{p} \in \tilde{F}_c$, $\mathbf{H}_h$ is negative definite on $\tilde{F}_c$. Note that $\tilde{F}_c$ is a convex set because for any $\mathbf{p}, \mathbf{q} \in \tilde{F}_c$ and $\mathbf{r} = \lambda\mathbf{p} + (1 - \lambda)\mathbf{q}$ with any $\lambda \in (0, 1)$ we have

$$0 < r_i = \lambda p_i + (1 - \lambda)q_i < \lambda + (1 - \lambda) = 1, \ 1 \le i \le m$$

$$\sum_{i=1}^{m} r_i \hat{W}_i = \lambda \sum_{i=1}^{m} p_i \hat{W}_i + (1 - \lambda) \sum_{i=1}^{m} q_i \hat{W}_i = \lambda\mu_0 + (1 - \lambda)\mu_0 = \mu_0 \qquad (4.42)$$

$$\sum_{i=1}^{m} r_i = \lambda \sum_{i=1}^{m} p_i + (1 - \lambda) \sum_{i=1}^{m} q_i = \lambda + (1 - \lambda) = 1.$$

Thus, function $h(\mathbf{p})$ is strictly concave on $\tilde{F}_c$ by Theorem 3.3.8 of Bazarra, Sherali, and Shetty (1993; pages 93 and 79) [1]. To verify the conditions in Theorem 4.3.8 of Bazarra, Sherali, and Shetty (1993; page 164) [1], note that $X_p = \{\mathbf{p} \mid 0 < p_i < 1, 1 \le i \le m\}$ is a nonempty open set in $\mathbb{R}^>$, and that $h(\mathbf{p})$, $h_1(\mathbf{p}) = 1 - \sum_{i=1}^{m} p_i$, and $h_2(\mathbf{p}) = \mu_0 - \sum_{i=1}^{m} p_i \hat{W}_i$ are each from $\mathbb{R}^> \to \mathbb{R}$. Since $\hat{\mathbf{p}} \in X_p$ satisfies constraints $h_1(\mathbf{p}) = 0$ and $h_2(\mathbf{p}) = 0$, $\hat{\mathbf{p}}$ is a feasible solution for (4.19) (Bazarra, Sherali, and Shetty, 1993; page 99) [1]. Also, note that

with $v_1 = n(1 - \lambda_0\mu_0)$ and $v_2 = n\lambda_0$, the KKT conditions are satisfied because

$$
\begin{aligned}
&\nabla h(\hat{\mathbf{p}}) + v_1 \nabla h_1(\hat{\mathbf{p}}) + v_2 \nabla h_2(\hat{\mathbf{p}}) \\
&= \begin{pmatrix} nw_1/\hat{p}_1 \\ \vdots \\ nw_m/\hat{p}_m \end{pmatrix} + n(1 - \lambda_0\mu_0) \begin{pmatrix} -1 \\ \vdots \\ -1 \end{pmatrix} + n\lambda_0 \begin{pmatrix} -\hat{W}_1 \\ \vdots \\ -\hat{W}_m \end{pmatrix} \\
&= \begin{pmatrix} n + n\lambda_0\hat{W}_1 - n\lambda_0\mu_0 \\ \vdots \\ n + n\lambda_0\hat{W}_m - n\lambda_0\mu_0 \end{pmatrix} + \begin{pmatrix} -n + n\lambda_0\mu_0 - n\lambda_0\hat{W}_1 \\ \vdots \\ -n + n\lambda_0\mu_0 - n\lambda_0\hat{W}_m \end{pmatrix} = \mathbf{0}.
\end{aligned}
$$

Since $h(\hat{\mathbf{p}})$ is concave and differentiable on $\tilde{F}_c$, $h(\hat{\mathbf{p}})$ is psuedoconcave on $\tilde{F}_c$ (Bazarra, Sherali, and Shetty, 1993; page 116). [1] Note that both h_1 and h_2 are linear functions, which means that both h_1 and h_2 are quasiconvex and quasiconcave on $\tilde{F}_c$ (Bazarra, Sherali, and Shetty, 1993; pages 116 and 118). [1] Thus, by Theorems 3.4.2 and 4.3.8 of Bazarra, Sherali, and Shetty (1993; pages 101 and 164), [1] $\hat{\mathbf{p}}$ is the unique optimal solution to (4.19), which complets the proof.

From (4.20) in Lemma 4.2, R_0 in (4.16) can be rewritten as

$$
R_0 = R(\mathbf{O}_n) = \prod_{i=1}^{m} \left(\frac{\tilde{p}_i}{w_i}\right)^{nw_i} = \prod_{i=1}^{m} \left(\frac{1}{1 + \lambda_0(\hat{W}_i - \mu_0)}\right)^{nw_i} \tag{4.43}
$$

and the rejection region for (4.7) analogous to (1.35) is given by

$$
\begin{aligned}
\{\mathbf{O}_n \mid R(\mathbf{O}_n) \leq c\} &= \left\{\mathbf{O}_n \,\middle|\, \prod_{i=1}^{m} \left(\frac{\tilde{p}_i}{w_i}\right)^{nw_i} \leq c\right\} \\
&= \left\{\mathbf{O}_n \,\middle|\, \prod_{i=1}^{m} \left(\frac{1}{1 + \lambda_0(\hat{W}_i - \mu_0)}\right)^{nw_i} \leq c\right\},
\end{aligned} \tag{4.44}
$$

for some predetermined $0 < c < 1$. For level of significance $0 < \alpha < 1$, we have:

$$\alpha = P\{\text{reject } H_0 \mid H_0 \text{ true}\} = P\{R_0 \le c \mid H_0\} = P\{-2\log R_0 \ge -2\log c \mid H_0\}. \quad (4.45)$$

The following theorem establishes the limiting distribution of $-2\log R_0$ under H_0 in (4.7).

Let

$$0 < \mu_X = \int x dF_X(x) < \infty, \qquad 0 < \sigma^2 = \int (x - \mu_X)^2 dF_X(x) < \infty \quad (4.46)$$

$$\hat{\mu} = \int x d\hat{F}_n(x), \qquad \sigma_n^2 = \int (x - \hat{\mu})^2 d\hat{F}_n(x) = \sum_{i=1}^{m} w_i(\hat{W}_i - \hat{\mu})^2 \quad (4.47)$$

Theorem 4.3. *Assume H_0 in hypothesis test (4.7) holds and assume:*

(AS8) $\quad E|V^X|^3 < \infty$ *and* $E|V^Y|^3 < \infty$;

(AS9) $\quad \sqrt{n} \int x d(\hat{F}_n(x; \hat{\gamma}) - F_X(x)) \xrightarrow{D} N(0, \tau^2), \quad as\ n \to \infty$;

(AS10) $\quad \int x^3 d\hat{F}_n(x; \hat{\gamma}) \xrightarrow{P} \int x^3 dF_X(x), \quad as\ n \to \infty.$

Then, we have:

$$-2\log R_0 \xrightarrow{D} \tau^2/\sigma^2 \chi_{(1)}^2, \quad as\ n \to \infty, \quad (4.48)$$

where $\chi_{(1)}^2$ represents a chi-squared random variable with 1 degree of freedom.

Remark 4.2. Note that (AS9) and (AS10) may be expected based on the asymptotic normality of estimator $\hat{\gamma}$ and the uniform consistency of $\hat{F}_n$. For assumption (AS8), we introduce the notation V^X and V^Y, as follows. For right censored data, doubly censored data, interval censored data Case 1 and partly-interval censored case 1 data given by (1.16)-(1.18) and (1.21), we have $\mathbf{O}_i^X = (V_i^X, \delta_i)$, where $V_i^X = \min\{X_i, C_i^X\}$ for right censored data; $V_i^X = \max\{\min\{X_i, C_i^X\}, D_i^X\}$ for doubly censored data; $V_i^X = C_i^X$ for interval censored

data Case 1 ; and $V_i^X = X_i\Delta_i + C_i^X(1-\Delta_i)$ with $P\{\Delta_i = 1\} = \rho_0^X > 0$, $P\{\Delta_i = 0\} = \rho_1^X > 0$, and $\rho_0^X + \rho_1^X = 1$ for partly-interval censored data Case 1. Since for these types of censored data (X_i, C_i^X) or (X_i, C_i^X, D_i^X), or C_i^X or (X_i, C_i^X, Δ_i) are $i.i.d.$, we know that V_i^X are $i.i.d.$ random variables. In the case of interval censored data Case 2 given by (1.19), we have $\mathbf{O}_i^X = (\mathbf{V}_i^X, \delta_i)$ with $\mathbf{V}_i^X = (C_i^X, D_i^X)$ and we denote

$$|V_i^X|^3 \equiv |C_i^X|^3 + |D_i^X|^3, \tag{4.49}$$

where C_i^X and D_i^X are $i.i.d.$, respectively.

Before proving Theorem 4.3, we state and prove the following lemma.

Lemma 4.4. *We have*

$$\max_{1 \leq i \leq m} |\hat{W}_i - \mu_0| = O_p(n^{1/3}). \tag{4.50}$$

Proof To establish (4.50) for all types of censored data, we note that

$$\max_{1 \leq i \leq m} |\hat{W}_i - \mu_0| \leq \max_{1 \leq i \leq m} |\hat{W}_i| + |\mu_0| \leq \max_{1 \leq i \leq m_1} |W_i^X| + |\hat{\gamma}| \max_{1 \leq i \leq m_0} |W_i^Y| + |\mu_0|. \tag{4.51}$$

For brevity, we discuss $\mathbf{O}_i^X$'s because $\mathbf{O}_j^Y$'s can be handled similarly. With notation V_i^X defined in Remark 4.1, note that

$$V_1^X, \ldots, V_{n_1}^X \text{ are } i.i.d. \text{ with } E|V^X|^3 < \infty \quad \Rightarrow \quad \max_{1 \leq i \leq n_1} |V_i^X| = O_p(n_1^{1/3}) \tag{4.52}$$

because

$$P\{\max_{1\leq i\leq n_1}|V_i^X| > n_1^{1/3}\} = 1 - \left[P\{|V_i^X| \leq n_1^{1/3}\}\right]^{n_1}$$

$$= \left(1 + [P\{|V_i^X| \leq n_1\}]^2 + \ldots + [P\{|V_i^X| \leq n_1\}]^{n_1-1}\right)\left(1 - P\{|V_i^X| \leq n_1^{1/3}\}\right)$$

$$\leq n_1[1 - P\{|V_i^X| \leq n_1^{1/3}\}] = n_1[P\{|V_i^X| > n_1^{1/3}\}] \leq n_1 \int_{|v|^3 > n_1} 1 dF_V^X(v)$$

$$\leq n_1 \int_{|v|^3 > n_1} \frac{|v|^3}{n_1} dF_V^X(v) = \int |v|^3 I\{|v|^3 > n_1\} = ^{n_1 \to \infty} 0.$$

Then, for all the types of censored data aforementioned, from (4.52), assumption (AS9) in

Theorem 4.3 and the fact

$$\{W_1^X, \ldots, W_{m_1}^X\} \subset \{V_1^X, \ldots, V_{n_1}^X\} \text{ or } \{C_1^X, D_1^X, \ldots, C_{n_1}^X, D_{n_1}^X\}, \tag{4.53}$$

we have

$$\max_{1\leq i\leq m_1}|W_i^X| \leq \max_{1\leq i\leq n_1}|V_i^X| = O_p(n_1^{1/3}) \tag{4.54}$$

or

$$\max_{1\leq i\leq m_1}|W_i^X| \leq \max_{1\leq i\leq n_1}|C_i^X| + \max_{1\leq i\leq n_1}|D_i^X| = O_p(n_1^{1/3}). \tag{4.55}$$

Similarly, for all the types of censored data aforementioned, from (4.3)

$$\max_{1\leq i\leq m_0}|W_i^Y| = O_p(n_0^{1/3}). \tag{4.56}$$

Since $\hat{\gamma} \xrightarrow{P} \gamma_0 \Rightarrow \hat{\gamma} = O_p(1)$, we know that (4.50) follows from (4.51), (4.54), (4.55), and

(4.56), which completes the proof.

Proof of Theorem 4.3 From (4.43) and (4.20) we can write an expression for $\log R_0$ as follows

$$
\begin{aligned}
\log R_0 &= \log \prod_{i=1}^{m} \left(\frac{\tilde{p}_i}{w_i} \right)^{nw_i} = n \sum_{i=1}^{m} w_i \log \left[\frac{w_i}{1 + \lambda_0(\hat{W}_i - \mu_0)} \right] - n \sum_{i=1}^{m} w_i \log w_i \\
&= -n \sum_{i=1}^{m} w_i \log(1 + \lambda_0(\hat{W}_i - \mu_0)).
\end{aligned}
\tag{4.57}
$$

To determine the asymptotic behavior of $-2 \log R_0$, we need to study the asymptotic behavior of λ_0. First, we note that from (4.14), $\hat{\mu}$ and σ_n^2 in (4.47) can be written as follows:

$$
\hat{\mu} = \int x d\hat{F}_n(x) = \rho_0 \hat{\mu}_X + \hat{\gamma} \rho_1 \hat{\mu}_Y = \rho_0 \sum_{i=1}^{m_0} \hat{p}_i^X W_i^X + \rho_1 \sum_{i=1}^{m_1} \hat{p}_i^Y (\hat{\gamma} W_i^Y) = \sum_{i=1}^{m} w_i \hat{W}_i \tag{4.58}
$$

and

$$
\sigma_n^2 = \int (x - \hat{\mu})^2 d\hat{F}_n(x) = \sum_{i=1}^{m} w_i (\hat{W}_i - \hat{\mu})^2. \tag{4.59}
$$

Denote $S^2 = \int (x - \mu_0)^2 d\hat{F}_n(x)$. Then,

$$
\begin{aligned}
S^2 &= \sum_{i=1}^{m} w_i (\hat{W}_i - \mu_0)^2 = \sum_{i=1}^{m} w_i (\hat{W}_i - \hat{\mu} + \hat{\mu} - \mu_0)^2 \\
&= \sum_{i=1}^{m} w_i (\hat{W}_i - \hat{\mu})^2 + 2 \sum_{i=1}^{m} w_i (\hat{W}_i - \hat{\mu})(\hat{\mu} - \mu_0) + \sum_{i=1}^{m} w_i (\hat{\mu} - \mu_0)^2 \\
&\overset{(4.58)}{=\!=\!=} \sum_{i=1}^{m} w_i (\hat{W}_i - \hat{\mu})^2 + (\hat{\mu} - \mu_0)^2 = \sigma_n^2 + O_p(n^{-1}).
\end{aligned}
\tag{4.60}
$$

Since

$$
\sigma_n^2 = \int (x - \hat{\mu})^2 d\hat{F}_n(x) \overset{P}{\to} \sigma^2 = \int (x - \mu_X) dF(x) \tag{4.61}
$$

due to $|\hat{F}_n(x) - F(x)| \overset{a.s.}{\to} 0$ and Theorem 4.1 we have

$$
S^2 \overset{P}{\to} \sigma^2, \quad \text{as } n \to \infty, \tag{4.62}
$$

which implies $S^2 > 0$ in probability, as $n \to \infty$.

Now we show that for any $1/3 < q < 1/2$, we have

$$\lambda_0 = O_p(n^{-q}). \tag{4.63}$$

From (4.50) we have in probability

$$1 + n^{-q}(\hat{W}_i - \mu_0) \le 1 + n^{-q}|\hat{W}_i - \mu_0| \le 1 + n^{-q+1/3}$$

and in (4.80), we have

$$
\begin{aligned}
g(n^{-q}) &= \sum_{i=1}^{m} \frac{w_i(\hat{W}_i - \mu_0)}{1 + n^{-q}(\hat{W}_i - \mu_0)} \\
&= \sum_{i=1}^{m} \frac{w_i(\hat{W}_i - \mu_0)[1 + n^{-q}(\hat{W}_i - \mu_0)] - n^{-q}w_i(\hat{W}_i - \mu_0)^2}{1 + n^{-q}(\hat{W}_i - \mu_0)} \\
&= \sum_{i=1}^{m} w_i(\hat{W}_i - \mu_0) - n^{-q} \sum_{i=1}^{m} \frac{w_i(\hat{W}_i - \mu_0)^2}{1 + n^{-q}(\hat{W}_i - \mu_0)} \\
&= \hat{\mu} - \mu_0 - n^{-q} \sum_{i=1}^{m} \frac{w_i(\hat{W}_i - \mu_0)^2}{1 + n^{-q}(\hat{W}_i - \mu_0)} \\
&\le \hat{\mu} - \mu_0 - n^{-q} \sum_{i=1}^{m} \frac{w_i(\hat{W}_i - \mu_0)^2}{1 + n^{-q+1/3}} = \hat{\mu} - \mu_0 - \frac{n^{-q}S^2}{1 + n^{-q+1/3}} \\
&= \hat{\mu} - \mu_0 - \frac{S^2}{n^q + n^{1/3}}.
\end{aligned}
$$

In turn, by assumption (AS9) and (4.62), we have

$$P\{g(n^{-q}) \ge 0\} = P\{n^{1/2}g(n^{-q}) \ge 0\} \le P\left\{n^{1/2}(\hat{\mu} - \mu_0) - \frac{n^{1/2}S^2}{n^q + n^{1/3}} \ge 0\right\}$$

$$= P\left\{O_p(1) - \frac{n^{1/2}S^2}{n^q + n^{1/3}} \ge 0\right\} \xrightarrow{n \to \infty} 0$$

because $-\frac{n^{1/2}S^2}{n^q + n^{1/3}} \xrightarrow{P} -\infty$, as $n \to \infty$. Thus, we have that $g(n^{-q}) < 0$ in probability.

Similarly, we can show $g(-n^{-q}) > 0$ in probability. Since $g(\lambda)$ is strictly decreasing on J,

(4.63) follows because

$$g(n^{-q}) < 0 < g(-n^{-q}) \quad \Rightarrow \quad g(n^{-q}) < g(\lambda_0) < g(-n^{-q})$$

$$\Rightarrow \quad -n^{-q} < \lambda_0 < n^{-q} \quad \Rightarrow \quad |\lambda_0| < n^{-q}.$$

To get an asymptotic expression for λ_0, we let $h = g^{-1}$ and note that $h(0) = \lambda_0$ and $h(\hat{\mu} - \mu_0) = 0$ because $g(\lambda_0) = 0$ and $g(0) = \hat{\mu} - \mu_0$ (by (4.58)), respectively. From the Taylor Expansion of h, we obtain

$$\lambda_0 = h(0) = h(\hat{\mu} - \mu_0) + (0 - (\hat{\mu} - \mu_0))h'(\xi) = -(\hat{\mu} - \mu_0)h'(\xi), \qquad (4.64)$$

where $|\xi| \leq |\hat{\mu} - \mu_0|$. Thus, λ_0 can be written as

$$
\begin{aligned}
\lambda_0 &= \frac{-(\hat{\mu} - \mu_0)}{g'(\eta)} = \frac{-(\hat{\mu} - \mu_0)}{-\sum_{i=1}^{m} \frac{w_i(\hat{W}_i - \mu_0)^2}{(1+\eta(\hat{W}_i - \mu_0))^2}} \cdot \frac{\sum_{i=1}^{m} w_i(\hat{W}_i - \mu_0)^2}{S^2} \\
&= \frac{(\hat{\mu} - \mu_0)}{S^2} \cdot \frac{\sum_{i=1}^{m} w_i(\hat{W}_i - \mu_0)^2}{\sum_{i=1}^{m} \frac{w_i(\hat{W}_i - \mu_0)^2}{(1+\eta(\hat{W}_i - \mu_0))^2}} = \frac{(\hat{\mu} - \mu_0)}{S^2} \cdot r_0,
\end{aligned}
\qquad (4.65)
$$

where $\eta = g^{-1}(\xi)$ with $|\eta| \leq |\lambda_0|$ and

$$r_0 = \frac{\sum_{i=1}^{m} w_i(\hat{W}_i - \mu_0)^2}{\sum_{i=1}^{m} \frac{w_i(\hat{W}_i - \mu_0)^2}{(1+\eta(\hat{W}_i - \mu_0))^2}}. \qquad (4.66)$$

To examine the asymptotic property of r_0, we note that from (4.63) and $|\eta| \leq |\lambda_0|$, we have $\eta = O_p(n^{-q})$, in turn, (4.50) gives

$$|\eta| \max_{1 \leq i \leq m} |\hat{W}_i - \mu_0| = O_p(n^{-q})O_p(n^{1/3}) = O_p(n^{-q+1/3}) = o_p(1). \qquad (4.67)$$

Thus, we have

$$r_0 \xrightarrow{P} 1, \qquad (4.68)$$

102

because

$$|r_0 - 1| = \left| \frac{\sum_{i=1}^m w_i(\hat{W}_i - \mu_0)^2}{\sum_{i=1}^m \frac{w_i(\hat{W}_i - \mu_0)^2}{[1+\eta(\hat{W}_i - \mu_0)]^2}} - 1 \right| = \left| \frac{\sum_{i=1}^m w_i(\hat{W}_i - \mu_0)^2 - \sum_{i=1}^m \frac{w_i(\hat{W}_i - \mu_0)^2}{[1+\eta(\hat{W}_i - \mu_0)]^2}}{\sum_{i=1}^m \frac{w_i(\hat{W}_i - \mu_0)^2}{[1+\eta(\hat{W}_i - \mu_0)]^2}} \right|$$

$$= \left| \frac{\sum_{i=1}^m \frac{w_i(\hat{W}_i - \mu_0)^2[1+\eta(\hat{W}_i - \mu_0)]^2 - w_i(\hat{W}_i - \mu_0)^2}{[1+\eta(\hat{W}_i - \mu_0)]^2}}{\sum_{i=1}^m \frac{w_i(\hat{W}_i - \mu_0)^2}{[1+\eta(\hat{W}_i - \mu_0)]^2}} \right| = \left| \frac{\sum_{i=1}^m \frac{\eta^2 w_i(\hat{W}_i - \mu_0)^4 + 2\eta w_i(\hat{W}_i - \mu_0)^3}{[1+\eta(\hat{W}_i - \mu_0)]^2}}{\sum_{i=1}^m \frac{w_i(\hat{W}_i - \mu_0)^2}{[1+\eta(\hat{W}_i - \mu_0)]^2}} \right|$$

$$\leq \left| \frac{\sum_{i=1}^m \frac{|\eta| w_i(\hat{W}_i - \mu_0)^2[\max_{1\leq i\leq m}|\hat{W}_i - \mu_0|][2+|\eta|\max_{1\leq i\leq m}|\hat{W}_i - \mu_0|]}{[1+\eta(\hat{W}_i - \mu_0)]^2}}{\sum_{i=1}^m \frac{w_i(\hat{W}_i - \mu_0)^2}{[1+\eta(\hat{W}_i - \mu_0)]^2}} \right|$$

$$= \left[|\eta| \max_{1\leq i\leq m} |\hat{W}_i - \mu_0| \right] \left[2 + |\eta| \max_{1\leq i\leq m} |\hat{W}_i - \mu_0| \right]$$

$$\overset{(4.67)}{=\joinrel=} o_p(1)\Big(2 + o_p(1)\Big) \overset{P}{\longrightarrow} 0.$$

Hence, from (4.62), (4.3), (4.68) and Slutsky's Theorem part (ii) (Serfling, 1980; page 19), we have that as $n \to \infty$

$$\sqrt{n}\lambda_0 = r_0 \frac{\sqrt{n}(\hat{\mu} - \mu_0)}{S^2} \overset{D}{\to} N\left(0, \frac{\tau^2}{\sigma^4}\right) \quad \Rightarrow \quad \lambda_0 = O_p(n^{-1/2}). \tag{4.69}$$

Note that from assumption (AS10), we have:

$$\hat{\mu}_{3n} = \int (x - \hat{\mu})^3 d\hat{F}_n(x) = \sum_{i=1}^m w_i(\hat{W}_i - \hat{\mu})^3 \quad \text{and} \quad \mu_3 \equiv \int (x - \mu_0)^3 dF_X(x), \tag{4.70}$$

and we introduce the following notation:

$$\tilde{\mu}_{3n} \int (x - \mu_0)^3 d\hat{F}_n(x) = \sum_{i=1}^m w_i(\hat{W}_i - \mu_0)^3 = \sum_{i=1}^m w_i(\hat{W}_i - \hat{\mu} + \hat{\mu} - \mu_0)^3$$

$$= \sum_{i=1}^m w_i\left[(\hat{W}_i - \hat{\mu})^3 + (\hat{\mu} - \mu_0)^3\right] + 3\sum_{i=1}^m w_i\left[(\hat{W}_i - \hat{\mu})^2(\hat{\mu} - \mu_0) + (\hat{W}_i - \hat{\mu})(\hat{\mu} - \mu_0)^2\right]$$

$$\overset{(4.58)}{=\joinrel=} \hat{\mu}_{3n} + 3(\hat{\mu} - \mu_0)\hat{\sigma}_n^2 + (\hat{\mu} - \mu_0)^3 = \hat{\mu}_{3n} + 3(\hat{\mu} - \mu_0)\left[\hat{\sigma}_n^2 + (\hat{\mu} - \mu_0)^2\right]$$

$$\overset{(AS10)}{=\joinrel=} \hat{\mu}_{3n} + O_p(n^{-1/2}). \tag{4.71}$$

To study the asymptotic behavior of R_0, from (4.57), (4.58), (4.60), (4.71) and the Taylor Expansion of $\log(1+x)$, we write $-2\log R_0$ as follows:

$$-2\log R_0 = 2n\sum_{i=1}^{m} w_i \log\left[1+\lambda_0(\hat{W}_i - \mu_0)\right]$$

$$= 2n\sum_{i=1}^{m} w_i\left[\lambda_0(\hat{W}_i-\mu_0) - \frac{1}{2}\left[\lambda_0(\hat{W}_i-\mu_0)\right]^2 + \frac{1}{3}\left[\lambda_0(\hat{W}_i-\mu_0)\right]^3 - \frac{\left[\lambda_0(\hat{W}_i-\mu_0)\right]^4}{4(1+\xi_i)^4}\right]$$

$$= 2n\lambda_0\sum_{i=1}^{m} w_i(\hat{W}_i-\mu_0) - n\lambda_0^2\sum_{i=1}^{m} w_i(\hat{W}_i-\mu_0)^2 + \frac{2n\lambda_0^3}{3}\sum_{i=1}^{m} w_i(\hat{W}_i-\mu_0)^3$$

$$-\frac{n\lambda_0^4}{2}\sum_{i=1}^{m}\frac{w_i(\hat{W}_i-\mu_0)^4}{(1+\xi_i)^4}$$

$$= 2n\lambda_0(\hat{\mu}-\mu_0) - n\lambda_0^2 S^2 + \frac{2n\lambda_0^3}{3}\tilde{\mu}_{3n} - \frac{n\lambda_0^4}{2}\sum_{i=1}^{m}\frac{w_i(\hat{W}_i-\mu_0)^4}{(1+\xi_i)^4}, \qquad (4.72)$$

where ξ_i is between 0 and $\lambda_0(\hat{W}_i - \mu_0)$. Note that $|\xi_i| \leq |\lambda_0(\hat{W}_i - \mu_0)|$ for $1 \leq i \leq m$, and from (4.50) and (4.69),

$$\max_{1\leq i\leq m}|\lambda_0(\hat{W}_i-\mu_0)| = O_p(n^{-1/2})O_p(n^{1/3}) = o_p(1).$$

Thus, from (4.50), (4.60), and (4.69), we have that in probability

$$\lambda_0^4\sum_{i=1}^{m}\frac{w_i(\hat{W}_i-\mu_0)^4}{(1+\xi_i)^4} \leq \lambda_0^4\sum_{i=1}^{m}\frac{w_i(\hat{W}_i-\mu_0)^4}{(1/2)^4} = 16\lambda_0^4\sum_{i=1}^{m} w_i(\hat{W}_i-\mu_0)^4$$

$$\leq 16\lambda_0^4\max_{1\leq i\leq m}(\hat{W}_i-\mu_0)^2\sum_{i=1}^{m} w_i(\hat{W}_i-\mu_0)^2$$

$$= 16\,O_p(n^{-2})O_p(n^{2/3})S^2 = O_p(n^{-4/3}).$$

Hence, equation (4.72) can be written as

$$-2\log R_0 = 2n\lambda_0(\hat{\mu}-\mu_0) - n\lambda_0^2 S^2 + \frac{2n\lambda_0^3}{3}\tilde{\mu}_{3n} + o_p(1). \qquad (4.73)$$

104

From (4.3) and (4.71), we have $\tilde{\mu}_{3n} = O_p(1)$. From (4.62), we have $S^2 = O_p(1)$. From (4.3), we have $\hat{\mu} - \mu_0 = O_p(n^{-1/2})$. From (4.68), we have $2r_0 - r_0^2 \xrightarrow{P} 1$. Thus, from (4.65), we have in (4.73),

$$
\begin{aligned}
-2\log R_0 &= 2n\frac{(\hat{\mu} - \mu_0)^2}{S^2}r_0 - n\frac{(\hat{\mu} - \mu_0)^2}{S^2}r_0^2 + \frac{2n\tilde{\mu}_{3n}}{3}\frac{(\hat{\mu} - \mu_0)^3}{S^6}r_0^3 + o_p(1) \\
&= \left[\frac{\sqrt{n}(\hat{\mu} - \mu_0)}{S}\right]^2 \left[2r_0 - r_0^2 + \frac{2r_0^3\tilde{\mu}_{3n}(\hat{\mu} - \mu_0)}{3S^4}\right] + o_p(1) \\
&= \left[\frac{\sqrt{n}(\hat{\mu} - \mu_0)}{S}\right]^2 \left(1 + o_p(1) + O_p(n^{-1/2})\right) + o_p(1).
\end{aligned} \tag{4.74}
$$

From (4.3), (4.62), and Slutsky's Theorem part (ii) [38]*page 19, we have

$$
-2\log R_0 = \left(\frac{\sqrt{n}(\hat{\mu} - \mu_0)}{\tau}\right)^2 \frac{\tau^2}{S^2}\left(1 + o_p(1)\right) + o_p(1) \xrightarrow{D} \frac{\tau^2}{\sigma^2}\chi^2_{(1)}, \quad \text{as } n \to \infty,
$$

which completes the proof.

From Theorem 4.3, in (4.45) we have

$$
\alpha = P\left\{-2\log R_0 \geq -2\log c \mid H_0\right\} \approx P\left\{\chi^2_{(1)} \geq -2\log c(\sigma^2/\tau^2)\right\},
$$

where σ^2/τ^2 is a constant that needs to be estimated. One possible approach is to use the following bootstrap procedure:

Bootstrap procedure for estimating σ^2/τ^2 :

Step 1. Compute $\hat{s}^2$ and $\hat{\mu}_{3n}$.

Step 2. Generate bootstrap samples $\mathbf{O}_1^{X*}, \ldots, \mathbf{O}_{n_1}^{X*}$ and $\mathbf{O}_1^{Y*}, \ldots, \mathbf{O}_{n_0}^{Y*}$ as in (2.26).

Step 3. Compute $\hat{G}^*$ and $\hat{H}^*$ as in (2.2).

Step 4. Compute $\hat{F}_n^*(\cdot; \hat{\eta})$ as in (2.20).

Step 5. Estimate $-2\log R_0$.

Then, from the chi-squared distribution table, constant c is calibrated by

$$-2\log c/\hat{C}_1 = \chi^2_{(1),\alpha} \quad \Rightarrow \quad c = e^{-\hat{C}_1 \chi^2_{(1),\alpha}/2}, \tag{4.75}$$

where $\chi^2_{(1),\alpha}$ is defined as the value satisfying $P\{\chi^2_{(1)} > \chi^2_{(1),\alpha}\} = \alpha$. With c determined in (4.75), the rejection region for test (4.7) is given in (4.44).

4.3.2 Confidence Intervals

From (4.14) and R_0 in (4.16), the weighted empirical likelihood ratio confidence region analogous to (1.40) for $\mu_X \equiv \mu_0$ is given by:

$$S = \left\{ \int x dF(x) \ \middle| \ R(F;\hat{\gamma}) \geq c, \ F \ll F_n \right\} = \left\{ \sum_{i=1}^{m} p_i \hat{W}_i \ \middle| \ \mathbf{p} \in \tilde{E}_c \right\}, \tag{4.76}$$

where

$$\tilde{E}_c = \left\{ \mathbf{p} \ \middle| \ 0 \leq p_i \leq 1, 1 \leq i \leq m; \ \sum_{i=1}^{m} p_i = 1; \ \prod_{i=1}^{m} \left(\frac{p_i}{w_i} \right)^{nw_i} \geq c \right\}. \tag{4.77}$$

The following two lemmas establish properties of S in (4.76) and establish a relationship between S and R_0 in (4.16).

Lemma 4.5. S is an interval that satisfies $S = [X_L, X_U]$, where

$$X_L = \inf_{\mathbf{p} \in \tilde{E}_c} \sum_{i=1}^{m} p_i \hat{W}_i \quad and \quad X_U = \sup_{\mathbf{p} \in \tilde{E}_c} \sum_{i=1}^{m} p_i \hat{W}_i. \tag{4.78}$$

Proof First, we let $y \in S$, which implies that $y = \sum_{i=1}^{m} p_i^* \hat{W}_i$ for some $\mathbf{p}^* \in E_c$. Then, we have

$$X_L = \inf_{\mathbf{p} \in E_c} \sum_{i=1}^{m} p_i \hat{W}_i \leq y \leq \sup_{\mathbf{p} \in E_c} \sum_{i=1}^{m} p_i \hat{W}_i = X_U \quad \Rightarrow \quad y \in [X_L, X_U].$$

106

Now, we let $y \in [X_L, X_U]$. To show that $y \in S$, we first show that $\min_{\mathbf{p} \in E_c} \sum_{i=1}^{m} p_i \hat{W}_i$ and $\max_{\mathbf{p} \in E_c} \sum_{i=1}^{m} p_i \hat{W}_i$ are attained on the set E_c. Note that E_c is a subset of $\mathbb{R}^{>}$ and is bounded because of the constraint $0 \leq p_i \leq 1$, $i = 1, \ldots, m$. Also note that if $\mathbf{p}^{(k)} \to \mathbf{p}^0$, as $k \to \infty$, for a sequence $\mathbf{p}^{(k)} \in E_c$, we have

$$\prod_{i=1}^{m} \left(\frac{p_i^{(k)}}{w_i} \right)^{nw_i} \geq c, \ \sum_{i=1}^{m} p_i^{(k)} = 1, \ 0 \leq p_i^{(k)} \leq 1, \ \text{for } i = 1, \ldots, m,$$

and we have

$$c \leq \lim_{k \to \infty} \prod_{i=1}^{m} \left(\frac{p_i^{(k)}}{w_i} \right)^{nw_i} = \prod_{i=1}^{m} \left(\frac{\lim_{k \to \infty} p_i^{(k)}}{w_i} \right)^{nw_i} = \prod_{i=1}^{m} \left(\frac{p_i^0}{w_i} \right)^{nw_i}$$

$$1 = \lim_{k \to \infty} \sum_{i=1}^{m} p_i^{(k)} = \sum_{i=1}^{m} \lim_{k \to \infty} p_i^{(k)} = \sum_{i=1}^{m} p_i^0$$

$$0 \leq \lim_{k \to \infty} p_i^{(k)} = p_i^0 \leq 1, i = 1, \ldots, m,$$

which implies that $\mathbf{p}^0 \in E_c$. Thus, E_c is closed; in turn, we know that E_c is compact. Since function $f(\mathbf{p}) = \sum_{i=1}^{m} p_i \hat{W}_i$ is linear and continuous on E_c, $f(\mathbf{p})$ attains its minimum and maximum on E_c. Hence, we have $X_L = \sum_{i=1}^{m} p_i^L$ and $X_U = \sum_{i=1}^{m} p_i^U$ for some $\mathbf{p}^L, \mathbf{p}^U \in E_c$, which gives

$$f(\mathbf{p}^L) = X_L \leq y \leq X_U = f(\mathbf{p}^U).$$

Now, it suffices to show that $y = \sum_{i=1}^{m} p_i^* \hat{W}_i$ for some $\mathbf{p}^* \in E_c$. Consider

$$h(\lambda) = f((1 - \lambda)\mathbf{p}^L + \lambda \mathbf{p}^U), \ \ 0 \leq \lambda \leq 1.$$

Then, we have $h(0) = f(\mathbf{p}^L) \leq y \leq f(\mathbf{p}^U) = h(1)$, and we know that $h(\lambda)$ is continuous and differentiable for $0 \leq \lambda \leq 1$ because

$$h(\lambda) = \sum_{i=1}^{m} \left[(1 - \lambda)p_i^L + \lambda p_i^U \right] \hat{W}_i = (1 - \lambda) \sum_{i=1}^{m} p_i^L \hat{W}_i + \lambda \sum_{i=1}^{m} p_i^U \hat{W}_i$$

107

is a linear function of λ. From the Mean-Value Theorem in Calculus we know that there exists a $\lambda^* \in [0,1]$ such that $h(0) \le y = h(\lambda^*) \le h(1) \Rightarrow f(\mathbf{p}^L) \le y = f(\mathbf{p}^*) = \sum_{i=1}^{m} p_i^* \hat{W}_i \le f(\mathbf{p}^U)$, where $\mathbf{p}^* = (1 - \lambda^*)\mathbf{p}^L + \lambda^* \mathbf{p}^U$. We complete the proof by showing $\mathbf{p}^* \in E_c$.

Since $0 \le p_i^L \le 1$ and $0 \le p_i^U \le 1$, $1 \le i \le m$, we have

$$0 \le p_i^* = (1 - \lambda^*)p_i^L + \lambda^* p_i^U \le (1 - \lambda^*) + \lambda^* = 1, \ i = 1, \dots, m$$

and since $\sum_{i=1}^{m} p_i^L = 1$ and $\sum_{i=1}^{m} p_i^U = 1$, we have

$$\sum_{i=1}^{m} p_i^* = \sum_{i=1}^{m} \left[(1 - \lambda^*)p_i^L + \lambda^* p_i^U \right] = (1 - \lambda^*) \sum_{i=1}^{m} p_i^L + \lambda^* \sum_{i=1}^{m} p_i^U = (1 - \lambda^*) + \lambda^* = 1.$$

To show that $\prod_{i=1}^{m} (\frac{p_i^*}{w_i})^{nw_i} \ge c$, we consider

$$g_1(\mathbf{p}) = \log \prod_{i=1}^{m} \left(\frac{p_i}{w_i} \right)^{nw_i} = n \sum_{i=1}^{m} w_i \log p_i - n \sum_{i=1}^{m} w_i \log w_i.$$

Then, for any $\mathbf{p} \in S = \{\mathbf{p} \mid p_i > 0\}$, the gradient vector and Hessian matrix (Bazarra, Sherali, and Shetty, page 90) [1] of g_1 exist and are given by, respectively,

$$\nabla g_1(\mathbf{p}) = \left(\frac{nw_1}{p_1}, \dots, \frac{nw_m}{p_m} \right)$$

and, with h_{ij} the component in the ith row and jth column of the Hessian matrix,

$$h_{ij} = \frac{\partial^2 g_1(\mathbf{p})}{\partial p_i \partial p_j} = \begin{cases} -\dfrac{nw_i}{p_i^2} & \text{if } i = j \\ 0 & \text{if } i \ne j. \end{cases} \implies \mathbf{H}_{g_1} = \text{diag} \left\{ -\frac{nw_1}{p_1^2}, \dots, -\frac{nw_m}{p_m^2} \right\}.$$

Since $\mathbf{H}_{g_1}$ is a diagonal matrix with diagonal elements $-\frac{nw_i}{p_i^2} < 0$, $1 \le i \le m$ for $p_i > 0$, $\mathbf{H}_{g_1}$ is negative definite on S. Note that S is a convex set (Bazarra, Sherali, and Shetty, 1993, page 34) [1] and thus, function $g_1(\mathbf{p})$ is strictly concave on S by Theorem 3.3.8 of (Bazarra,

Sherali, and Shetty, 1993, page 92). [1] Also note that $w_i > 0$ and $\mathbf{p}^L, \mathbf{p}^U \in E_c$, imply

$\mathbf{p}^L, \mathbf{p}^U \in S$ because $\prod_{i=1}^{m} \left(\frac{p_i^L}{w_i} \right)^{nw_i} \geq c > 0 \Rightarrow p_i^L > 0$, $1 \leq i \leq m$. Since g_1 is a strictly

concave function on S, from $g_1(\mathbf{p}^L) = \log \left[\prod_{i=1}^{m} \left(\frac{p_i^L}{w_i} \right)^{nw_i} \right] \geq \log(c)$ and $g_1(\mathbf{p}^U) \geq \log(c)$, we

have

$$
\begin{aligned}
\log \left[\prod_{i=1}^{m} \left(\frac{p_i^*}{w_i} \right)^{nw_i} \right] &= g_1(\mathbf{p}^*) = g_1((1 - \lambda^*)\mathbf{p}^L + \lambda^* \mathbf{p}^U) \\
&\geq (1 - \lambda^*)g_1(\mathbf{p}^L) + \lambda^* g_1(\mathbf{p}^U) \\
&\geq (1 - \lambda^*)\log(c) + \lambda^* \log(c) = \log(c),
\end{aligned}
$$

which implies that $\prod_{i=1}^{m} \left(\frac{p_i^*}{w_i} \right)^{nw_i} \geq c$, which completes the proof.

Lemma 4.6. $X_L \leq \mu_0 \leq X_U$ *if and only if* $R_0 \geq c$.

Proof Suppose $X_L \leq \mu_0 \leq X_U$. Then, we have $\mu_0 = \sum_{i=1}^{m} p_i^* \hat{W}_i$ for some $\mathbf{p}^* \in E_c$. Since

$\mathbf{p}^* \in E_c$, we have

$$
\prod_{i=1}^{m} \left(\frac{p_i^*}{w_i} \right)^{nw_i} \geq c, \ \sum_{i=1}^{m} p_i^* = 1, \ 0 \leq p_i^* \leq 1, \text{ for } i = 1, \ldots, m.
$$

Then, $\mathbf{p}^* \in F_c$ and we have

$$
c \leq \prod_{i=1}^{m} \left(\frac{p_i^*}{w_i} \right)^{nw_i} \leq \sup_{\mathbf{p} \in F_c} \prod_{i=1}^{m} \left(\frac{p_i}{w_i} \right)^{nw_i} = R_0.
$$

Now suppose $R_0 \geq c$. To show that $X_L \leq \mu_0 \leq X_U$, we first show that $\max_{\mathbf{p} \in F_c} \prod_{i=1}^{m} \left(\frac{p_i}{w_i} \right)^{nw_i}$

is attained on the set F_c. Note that F_c is a subset of $\mathbb{R}^{>}$ and is bounded because of the con-

straint $0 \leq p_i \leq 1$, $i = 1, \ldots, m$. Also note that if $\mathbf{p}^{(k)} \to \mathbf{p}^0$, as $k \to \infty$ for a sequence

$\mathbf{p}^{(k)} \in F_c$, we have

$$\sum_{i=1}^{m} p_i^{(k)} \hat{W}_i = \mu_0, \ \sum_{i=1}^{m} p_i^{(k)} = 1, \ 0 \le p_i^{(k)} \le 1, \ \text{for} \ i = 1, \dots, m,$$

and we have

$$\mu_0 = \lim_{k \to \infty} \sum_{i=1}^{m} p_i^{(k)} \hat{W}_i = \sum_{i=1}^{m} \lim_{k \to \infty} p_i^{(k)} \hat{W}_i = \sum_{i=1}^{m} p_i^0 \hat{W}_i$$

$$1 = \lim_{k \to \infty} \sum_{i=1}^{m} p_i^{(k)} = \sum_{i=1}^{m} \lim_{k \to \infty} p_i^{(k)} = \sum_{i=1}^{m} p_i^0$$

$$0 \le \lim_{k \to \infty} p_i^{(k)} = p_i^0 \le 1, i = 1, \dots, m,$$

which implies that $\mathbf{p}^0 \in F_c$. Thus, F_c is closed; in turn, we know that F_c is compact. Since

function $f(\mathbf{p}) = \prod_{i=1}^{m} (\frac{p_i}{w_i})^{nw_i}$ is continuous on F_c, $f(\mathbf{p})$ attains its maximum for some $\mathbf{p}^*$ on

F_c. Hence, we have

$$c \le R_0 = \sup_{\mathbf{p} \in F_c} \prod_{i=1}^{m} \left(\frac{p_i}{w_i} \right)^{nw_i} = \max_{\mathbf{p} \in F_c} \prod_{i=1}^{m} \left(\frac{p_i}{w_i} \right)^{nw_i} = \prod_{i=1}^{m} \left(\frac{p_i^*}{w_i} \right)^{nw_i}.$$

Since $\mathbf{p}^* \in F_c$, we have

$$\prod_{i=1}^{m} \left(\frac{p_i^*}{w_i} \right)^{nw_i} \ge c, \ \sum_{i=1}^{m} p_i^* \hat{W}_i = \mu_0, \ \sum_{i=1}^{m} p_i^* = 1, \ 0 \le p_i^* \le 1, \ \text{for} \ i = 1, \dots, m,$$

Hence, $\mu_0 = \sum_{i=1}^{m} p_i^* \hat{W}_i$ with $\mathbf{p}^* \in E_c$ which implies that $\mu_0 \in \left\{ \sum_{i=1}^{m} p_i \hat{W}_i \mid \mathbf{p} \in E_c \right\}$; in

turn, $\mu_0 \in [X_L, X_U]$, which completes the proof.

From Theorem 4.3 and Lemma 4.6, we have

$$P\left\{ X_L \le \mu_0 \le X_U \right\} = P\left\{ -2 \log R_0 \le -2 \log c \right\} \approx P\left\{ \chi_{(1)}^2 \le \frac{-2 \log c}{\hat{C}_1} \right\} = 1 - \alpha,$$

where c is determined by (4.75). The procedure for obtaining $\hat{C}_1$ is provided in Chapter 5

where we present the estimation algorithm.

4.3.3   Computation of Confidence Intervals

In this section, we discuss computation of X_L and X_U in (4.78). From (4.78), X_L and X_U are obtained, respectively, by solving

$$
\begin{cases}
\text{Minimize/Maximize} \quad f(\mathbf{p}) = \sum_{i=1}^{m} p_i \hat{W}_i \\[2mm]
\text{subject to:} \qquad 0 < p_i < 1, 1 \le i \le m; \ \sum_{i=1}^{m} p_i = 1; \ \prod_{i=1}^{m} \left(\frac{p_i}{w_i} \right)^{n w_i} \ge c.
\end{cases}
\tag{4.79}
$$

To find a candidate for a solution to (4.79) using the Lagrange multipliers, we denote

$$
G(\mathbf{p}, \beta, \lambda) = \sum_{i=1}^{m} p_i \hat{W}_i + \beta \left[\log c - n \sum_{i=1}^{m} w_i \log \left(\frac{p_i}{w_i} \right) \right] + \lambda \left[1 - \sum_{i=1}^{m} p_i \right]
\tag{4.80}
$$

then we have for $1 \le i \le m$

$$
0 = \frac{\partial G}{\partial p_i} = \hat{W}_i - \frac{n \beta w_i}{p_i} - \lambda \quad \Rightarrow \quad p_i = \frac{n \beta w_i}{\hat{W}_i - \lambda}.
\tag{4.81}
$$

Note that $n w_i > 0$ for $1 \le i \le m$ and to ensure that the denominator in (4.81) is not equal to 0, we need either $\lambda < \hat{W}_{(1)}$ or $\lambda > \hat{W}_{(m)}$, for which we have

$$
\text{when } \lambda < \hat{W}_{(1)}, \text{ we need } \beta > 0 \text{ in (4.81);}
$$
$$
\text{when } \lambda > \hat{W}_{(m)}, \text{ we need } \beta < 0 \text{ in (4.81).}
\tag{4.82}
$$

From (2.3), (4.14), (4.81), and the second constraint in (4.79), we have

$$0 = p_i\hat{W}_i - p_i\lambda - n\beta w_i \quad \Rightarrow \quad 0 = \sum_{i=1}^{m}\left[p_i\hat{W}_i - n\beta w_i\right] - \lambda$$

$$\Rightarrow \quad 0 = \sum_{i=1}^{m}\left[\frac{n\beta w_i}{\hat{W}_i - \lambda}\hat{W}_i - n\beta w_i\right] - \lambda$$

$$\Rightarrow \quad 0 = n\beta\sum_{i=1}^{m}\left[\frac{w_i\hat{W}_i}{\hat{W}_i - \lambda} - w_i\right] - \lambda$$

$$\Rightarrow \quad \lambda = n\beta\sum_{i=1}^{m}\frac{\lambda w_i}{\hat{W}_i - \lambda}$$

$$\Rightarrow \quad \beta = \frac{1}{n\sum_{i=1}^{m}\frac{w_i}{\hat{W}_i - \lambda}}. \tag{4.83}$$

From (4.82) and (4.83), we have

$$\text{when } \lambda < \hat{W}_{(1)}, \text{ we have } \beta > 0 \text{ in (4.83);} \tag{4.84}$$

$$\text{when } \lambda > \hat{W}_{(m)}, \text{ we have } \beta < 0 \text{ in (4.83).} \tag{4.85}$$

Substituting (4.83) into (4.81), we obtain for $\lambda < \hat{W}_{(1)}$ or $\lambda > \hat{W}_{(m)}$,

$$p_i = \frac{w_i}{(\hat{W}_i - \lambda)\sum_{j=1}^{m}\frac{w_j}{\hat{W}_j - \lambda}}, \quad 1 \le i \le m. \tag{4.86}$$

From equation (4.80) we have for $\lambda < \hat{W}_{(1)}$ or $\lambda > \hat{W}_{(m)}$,

$$0 = \frac{\partial G}{\partial \beta} = \log c - n\sum_{i=1}^{m}w_i\log\left(\frac{p_i}{w_i}\right)$$

$$= \log c - n\sum_{i=1}^{m}w_i\log\left(\frac{w_i}{w_i\left[\left(\hat{W}_i - \lambda\right)\sum_{j=1}^{m}\frac{w_j}{\hat{W}_j - \lambda}\right]}\right) \tag{4.87}$$

$$= \log c + n\sum_{i=1}^{m}w_i\log\left[\left(\hat{W}_i - \lambda\right)\sum_{j=1}^{m}\frac{w_j}{\hat{W}_j - \lambda}\right] \equiv g(\lambda),$$

where $g(\lambda)$ is well-defined on $(-\infty, \hat{W}_{(1)})$ and $(\hat{W}_{(m)}, \infty)$.

112

The following lemma provides expressions for the solutions of optimization problem (4.79).

Lemma 4.7. X_L and X_U given by

$$X_L = \left(\sum_{j=1}^{m} \frac{w_j}{\hat{W}_j - \lambda_L} \right)^{-1} \sum_{i=1}^{m} \left(\frac{w_i \hat{W}_i}{\hat{W}_i - \lambda_L} \right) \tag{4.88}$$

with λ_L as the unique solution of $g(\lambda)$ on $(-\infty, \hat{W}_{(1)})$ and

$$X_U = \left(\sum_{j=1}^{m} \frac{w_j}{\hat{W}_j - \lambda_U} \right)^{-1} \sum_{i=1}^{m} \left(\frac{w_i \hat{W}_i}{\hat{W}_i - \lambda_U} \right) \tag{4.89}$$

with λ_L as the unique solution of $g(\lambda)$ on $(\hat{W}_{(m)}, \infty)$ are the unique minimum and maximum solutions for (4.79), respectively.

Proof We first note that in the proof of Lemma 4.6 it is shown that the minimum $\hat{p}_i^L$'s and maximum $\hat{p}_i^U$'s for X_L and X_U, respectively, are attained on the set E_c. Note that $nw_i > 0$, $1 \le i \le m$ implies $\hat{p}_i^L > 0$, $1 \le i \le m$ and similarly $\hat{p}_i^U > 0$, $1 \le i \le m$. Also, note that $\hat{p}_j^L = 1$ for some $1 \le j \le m$ implies that $\hat{p}_i^L = 0$ for all $i \ne j$ because of $\sum_{i=1}^{m} \hat{p}_i^L = 1$ and similarly, $\hat{p}_j^U = 1$ for some $1 \le j \le m$ implies that $\hat{p}_i^U = 0$ for all $i \ne j$ because of $\sum_{i=1}^{m} \hat{p}_i^U = 1$. Thus, the minimum and maximum solutions of (4.79) are attained on the set E_c^* where

$$E_c^* = \left\{ \mathbf{p} \ \middle| \ 0 < p_i < 1, \ 1 \le i \le m; \ \prod_{i=1}^{m} \left(\frac{p_i}{w_i} \right)^{nw_i} \ge c; \ \sum_{i=1}^{m} p_i = 1 \right\}. \tag{4.90}$$

113

Then, optimization problem (4.79) is equivalent to the following optimization problem:

$$\begin{cases} \text{Minimize/Maximize} \quad f(\mathbf{p}) = \sum_{i=1}^{m} p_i \hat{W}_i \\[2ex] \text{subject to:} \qquad g_1(\mathbf{p}) = \log c - n \sum_{i=1}^{m} w_i \log\left(\frac{p_i}{w_i}\right) \leq 0 \\[2ex] \qquad\qquad h_1(\mathbf{p}) = 1 - \sum_{i=1}^{m} p_i = 0 \\[2ex] \qquad\qquad 0 < p_i < 1, \ 1 \leq i \leq m. \end{cases} \tag{4.91}$$

First, based on (4.86), we define the notation $\hat{\mathbf{p}}^L = (\hat{p}_1^L, \ldots, \hat{p}_m^L)$ where

$$\hat{p}_i^L = \frac{w_i}{(\hat{W}_i - \lambda_L) \sum_{j=1}^{m} \frac{w_j}{\hat{W}_j - \lambda_L}}, \quad 1 \leq i \leq m, \tag{4.92}$$

and $\hat{\mathbf{p}}^U = (\hat{p}_1^U, \ldots, \hat{p}_m^U)$ where

$$\hat{p}_i^U = \frac{w_i}{(\hat{W}_i - \lambda_U) \sum_{j=1}^{m} \frac{w_j}{\hat{W}_j - \lambda_U}}, \quad 1 \leq i \leq m. \tag{4.93}$$

Next, for optimization problem (4.91) we verify the KKT conditions in Theorem 4.3.8 of

Bazarra, Sherali, and Shetty (1993, page 164). [1] Note that $X = \{\mathbf{p} \mid 0 < p_i < 1, 1 \leq i \leq m\}$

is a nonempty open set in $\mathbb{R}^>$, and that $f(\mathbf{p})$, $g_1(\mathbf{p})$, and $h_1(\mathbf{p})$ in (4.91) are each from

$\mathbb{R}^> \to \mathbb{R}$. Since $\hat{\mathbf{p}}^L$, $\hat{\mathbf{p}}^U \in X$ satisfy constraints $g_1(\mathbf{p}) = 0$ and $h_1(\mathbf{p}) = 0$, $\hat{\mathbf{p}}^L$ and $\hat{\mathbf{p}}^U$ are

both feasible solutions for optimization problem (4.91) (Bazarra, Sherali, and Shetty, 1993,

page 99). [1]

Since $f(\mathbf{p})$ is a linear function, and thus differentiable on E_c^*, it is both psuedoconvex and

psuedoconcave on E_c^* (Bazarra, Sherali, and Shetty, 1993, page 116 and 118). [1] Similarly,

since $h_1(\mathbf{p})$ is a linear function, and thus differentiable on E_c^*, it is both quasiconvex and

quasiconcave on E_c^* ((Bazarra, Sherali, and Shetty, 1993, page 116 and 118). [1] Also note

that $g_1(\mathbf{p})$ is both quasiconvex and quasiconcave on E_c^*.

Next, we consider the sign of the Lagrange multiplier β, given in (4.83). When $\beta > 0$,

the feasible solution $\hat{\mathbf{p}}^L$ is a candidate for the solution of the minimization problem in (4.91).

Note that the KKT conditions are satisfied for $\hat{\mathbf{p}}^L$ because

$$
\begin{aligned}
\boldsymbol{\nabla} f(\hat{\mathbf{p}}^L) + \beta \boldsymbol{\nabla} g_1(\hat{\mathbf{p}}^L) + \lambda_L \boldsymbol{\nabla} h_1(\hat{\mathbf{p}}^L) \;&=\; \begin{pmatrix} \hat{W}_1 \\ \vdots \\ \hat{W}_m \end{pmatrix} + \frac{1}{n \sum_{i=1}^m \frac{w_i}{\hat{W}_i - \lambda_L}} \begin{pmatrix} \frac{-nw_1}{\hat{p}_1^L} \\ \vdots \\ \frac{-nw_m}{\hat{p}_m^L} \end{pmatrix} + \lambda_L \begin{pmatrix} -1 \\ \vdots \\ -1 \end{pmatrix} \\[2em]
&=\; \begin{pmatrix} \hat{W}_1 \\ \vdots \\ \hat{W}_m \end{pmatrix} + \frac{1}{n \sum_{i=1}^m \frac{w_i}{\hat{W}_i - \lambda_L}} \begin{pmatrix} -n(\hat{W}_1 - \lambda_L) \sum_{i=1}^m \frac{w_i}{\hat{W}_i - \lambda_L} \\ \vdots \\ -n(\hat{W}_m - \lambda_L) \sum_{i=1}^m \frac{w_i}{\hat{W}_i - \lambda_L} \end{pmatrix} + \begin{pmatrix} -\lambda_L \\ \vdots \\ -\lambda_L \end{pmatrix} \\[2em]
&=\; \begin{pmatrix} \hat{W}_1 \\ \vdots \\ \hat{W}_m \end{pmatrix} + \begin{pmatrix} -(\hat{W}_1 - \lambda_L) \\ \vdots \\ -(\hat{W}_m - \lambda_L) \end{pmatrix} + \begin{pmatrix} -\lambda_L \\ \vdots \\ -\lambda_L \end{pmatrix} \;=\; \mathbf{0}.
\end{aligned}
$$

Thus, by Theorems 3.4.2 and 4.3.8 of (Bazarra, Sherali, and Shetty, 1993, pages 101 and

164), [1] $\hat{\mathbf{p}}^L$ is the unique solution to the minimization problem in (4.91).

Similarly, when $\beta < 0$, the feasible solution $\hat{\mathbf{p}}^U$ is a candidate for the solution of the

maximization problem in (4.91). Note that the KKT conditions are satisfied for $\hat{\mathbf{p}}^U$ because

$$
\begin{aligned}
\boldsymbol{\nabla} f(\hat{\mathbf{p}}^U) + \beta \boldsymbol{\nabla} g_1(\hat{\mathbf{p}}^U) + \lambda_U \boldsymbol{\nabla} h_1(\hat{\mathbf{p}}^U) \;=\;&
\begin{pmatrix} \hat{W}_1 \\ \vdots \\ \hat{W}_m \end{pmatrix}
+ \frac{1}{n \sum_{i=1}^{m} \frac{w_i}{\hat{W}_i - \lambda_U}}
\begin{pmatrix} \frac{-n w_1}{\hat{p}_1^U} \\ \vdots \\ \frac{-n w_m}{\hat{p}_m^U} \end{pmatrix}
+ \lambda_U \begin{pmatrix} -1 \\ \vdots \\ -1 \end{pmatrix} \\[2ex]
=\;&
\begin{pmatrix} \hat{W}_1 \\ \vdots \\ \hat{W}_m \end{pmatrix}
+ \frac{1}{n \sum_{i=1}^{m} \frac{w_i}{\hat{W}_i - \lambda_U}}
\begin{pmatrix} -n(\hat{W}_1 - \lambda_U) \sum_{i=1}^{m} \frac{w_i}{\hat{W}_i - \lambda_U} \\ \vdots \\ -n(\hat{W}_m - \lambda_U) \sum_{i=1}^{m} \frac{w_i}{\hat{W}_i - \lambda_U} \end{pmatrix}
+ \begin{pmatrix} -\lambda_U \\ \vdots \\ -\lambda_U \end{pmatrix} \\[2ex]
=\;&
\begin{pmatrix} \hat{W}_1 \\ \vdots \\ \hat{W}_m \end{pmatrix}
+ \begin{pmatrix} -(\hat{W}_1 - \lambda_U) \\ \vdots \\ -(\hat{W}_m - \lambda_U) \end{pmatrix}
+ \begin{pmatrix} -\lambda_U \\ \vdots \\ -\lambda_U \end{pmatrix}
\;=\; \mathbf{0}.
\end{aligned}
$$

Thus, by Theorems 3.4.2 and 4.3.8 of (Bazarra, Sherali, and Shetty, 1993, pages 101 and 164), [1] $\hat{\mathbf{p}}^U$ is the unique solution to the maximization problem in (4.91).

CHAPTER 5
SIMULATION STUDIES

In this chapter we discuss some relevant bootstrap procedures and provide simulation results

for the estimators for γ_0 and μ_X, given in Chapters 3−4, respectively. The organization of

this chapter is as follows. Section 5.1 reviews the Bootstrap Percentile Confidence Interval.

Section 5.2 gives simulation results on estimators for the scale parameter γ_0, discussed in

Chapter 3. Section 5.3 gives simulation results on estimators for μ_X, discussed in Chapter 4.

Section 5.4 gives simulation results on the treatment distribution estimator for F_X.

5.1   Review of Bootstrap Percentile Confidence Intervals

In this section, we discuss ideas of the bootstrap method and in particular, we outline the

main ideas of the bootstrap percentile confidence interval [12]. Let

$$X_1, \ldots, X_n \overset{i.i.d.}{\sim} F_0, \tag{5.1}$$

where F_0 is an unknown distribution function. Consider a parameter of interest given by

$\theta = T(\mathbf{X}; F)$, where $\mathbf{X} = (X_1, \ldots, X_n)$. Once a random sample is observed, we can obtain

an estimate $\hat{\theta}$ for θ using the plug-in principle as follows:

$$\hat{\theta} = T(\mathbf{X}; \hat{F}_n), \tag{5.2}$$

where $\hat{F}_n$ is the empirical distribution function that is calculated from the observed sample.

A *bootstrap sample* is a sample of size n drawn with replacement from the population of n objects $(X_1, \ldots, X_n)$ and is denoted as $\mathbf{X}^* = (X_1^*, \ldots, X_n^*)$.

The Bootstrap Percentile Confidence Interval can be computed as follows.

Bootstrap Percentile Confidence Interval Algorithm:

Step 1. Generate a bootstrap sample $X_1^* \ldots, X_n^*$ from the sample in (5.1).

Step 2. Calculate $\hat{\theta}^*(\mathbf{X}^*)$.

Step 3. Repeat Steps $1-3$ B times to obtain $\hat{\theta}^*(b)$, $b = 1, \ldots B$, where B is usually chosen

to be 1000.

Step 4. Arrange the $\hat{\theta}^*(i)'s$ in ascending order: $\hat{\theta}^*(1) \leq \cdots \leq \hat{\theta}^*(B)$

Step 5. Let $\hat{\theta}_B^{*(\alpha)}$ be the $100 \cdot \alpha$th percentile.

Then, a $(1 - \alpha) \cdot 100\%$ *Bootstrap Percentile Confidence Interval* is given by:

$$\left(\hat{\theta}_B^{*(\alpha/2)}, \hat{\theta}_B^{*(1-\alpha/2)} \right). \tag{5.3}$$

5.2   Point and Interval Estimators of a Scale Parameter

In this section, we provide simulation results on point estimates and interval estimates for scale parameter γ_0 in Two-Sample Accelerated Life Model (1.5). While only a few selected results are included in this dissertation, additional simulations yielded similar results. For

the simulations in this dissertation, we denote $\text{Exp}(\mu)$ as the exponential distribution with the mean μ.

5.2.1 Point Estimators for the Scale Parameter

For this section, we present simulation results for the point estimate $\hat{\gamma}_E$ in (3.2) and the point estimate $\hat{\gamma}_R$ as described in Steps 1-3 of the bootstrap procedure at the end of Subsection 3.2.3.

In Table 5.1, 10,000 right censored samples (1.16) of sizes 25, 50, and 100 are taken from $X \sim \text{Exp}(1)$ and $Y \sim \text{Exp}(2)$ with censoring distribution $C \sim \text{Exp}(4)$. Note that the censoring percentages for the X's and Y's are given in Table 5.1. In addition, for each method we provide the estimate, standard error (SE), error, and relative error.

From Tables 5.1 and 5.2, we see that rank based estimator $\hat{\gamma}_R$ has smaller relative errors for each of the sample sizes examined and thus, $\hat{\gamma}_R$ provides a better estimate for the true parameter γ_0 than the naive estimator $\hat{\gamma}_E$.

We also note that it appears that the rank based estimator $\hat{\gamma}_R$ tends to underestimate γ_0 while naive estimator $\hat{\gamma}_E$ tends to overestimate γ_0. This is something to be explored in future research.

Table 5.1: Point Estimators for the scale parameter γ_0

	$X \sim \text{Exp}(2), \ Y \sim \text{Exp}(1), \ C \sim \text{Exp}(4), \ \gamma_0 = 0.5$					
Sample Size	Censoring %	Method	Estimate	SE	Error	Relative Error
$n_X = n_Y = 25$	X: 20.10	$\hat{\gamma}_E$	0.5493	0.1882	+0.0493	+0.0987
	Y: 33.38	$\hat{\gamma}_R$	0.4853	0.1758	−0.0147	−0.0293
$n_X = n_Y = 50$	X: 20.02	$\hat{\gamma}_E$	0.5260	0.1256	+0.0260	+0.0519
	Y: 33.35	$\hat{\gamma}_R$	0.4909	0.1225	−0.0091	−0.0182
$n_X = n_Y = 100$	X: 20.01	$\hat{\gamma}_E$	0.5146	0.0885	+0.0146	+0.0292
	Y: 33.34	$\hat{\gamma}_R$	0.4939	0.0866	−0.0061	−0.0123

Table 5.2: Point Estimators for the scale parameter γ_0

	$X \sim \text{Exp}(2), \ Y \sim \text{Exp}(5), \ C \sim \text{Exp}(8), \ \gamma_0 = 0.4$					
Sample Size	Censoring %	Method	Estimate	SE	Error	Relative Error
$n_X = n_Y = 25$	X: 20.10	$\hat{\gamma}_E$	0.4488	0.1563	+0.0488	+0.1220
	Y: 38.48	$\hat{\gamma}_R$	0.3909	0.1433	−0.0091	−0.0226
$n_X = n_Y = 50$	X: 20.03	$\hat{\gamma}_E$	0.4270	0.1045	+0.0270	+0.0675
	Y: 38.46	$\hat{\gamma}_R$	0.3943	0.1001	−0.0057	−0.0142
$n_X = n_Y = 100$	X: 20.01	$\hat{\gamma}_E$	0.4157	0.0738	+0.0157	+0.0393
	Y: 38.49	$\hat{\gamma}_R$	0.3955	0.0701	−0.0045	−0.0113

Table 5.3: 90% C.I. for the scale parameter γ_0 with right censored exponential data

$X \sim \text{Exp}(1), \ Y \sim \text{Exp}(2), \ C \sim \text{Exp}(4), \ \gamma_0 = 0.5$			
Censoring % for X: 20.07%		Censoring % for Y: 33.48%	
$n_X = n_Y = 25$	Coverage	Mean Length of C.I.	s.d. Length of C.I.
Normal C.I. $(\hat{\gamma}_E)$ (3.12)	0.882	0.5811	0.2287
Normal C.I. $(\hat{\gamma}_R)$ (3.20)	0.825	0.5482	0.2146
Bootstrap Percentile C.I. $(\hat{\gamma}_E)$	0.866	0.5614	0.2200
Bootstrap Percentile C.I. $(\hat{\gamma}_R)$	0.838	0.5250	0.2064

5.2.2 Interval Estimators for the Scale Parameter

In Tables 5.3−5.8 we present the results for the interval estimates (3.12) and (3.20) respectively along with bootstrap percentile confidence intervals. For Tables 5.3, 5.4, and 5.5, 1,000 right censored samples (1.16) of sizes 25, 50, and 100 are generated from $X \sim \text{Exp}(1)$ and $Y \sim \text{Exp}(2)$ with censoring distribution $C \sim \text{Exp}(4)$. For Tables 5.6, 5.7, and 5.8, 1,000 right censored samples (1.16) of sizes 25, 50, and 100 are generated from $X \sim \text{Exp}(1)$ and $Y \sim \text{Exp}(2)$ with censoring distribution $C \sim \text{Exp}(4)$. For each tables, the interval estimates (3.12) and (3.20) are calculated. In addition, for each of the 1,000 simulation loops, we take 400 nested bootstrap samples and compute the bootstrap percentile confidence intervals for γ_0 based on $\hat{\gamma}_E$ and $\hat{\gamma}_R$, respectively as described in (5.3). Note that the censoring percentages for X and Y, respectively, are given as well.

Table 5.4: 90% C.I. for the scale parameter γ_0 with right censored exponential data

$X \sim \text{Exp}(1), \ Y \sim \text{Exp}(2), \ C \sim \text{Exp}(4), \ \gamma_0 = 0.5$			
Censoring % for X: 19.97%		Censoring % for Y: 33.33%	
$n_X = n_Y = 50$	Coverage	Mean Length of C.I.	s.d. Length of C.I.
Normal C.I. $(\hat{\gamma}_E)$ (3.12)	0.901	0.4008	0.1133
Normal C.I. $(\hat{\gamma}_R)$ (3.20)	0.871	0.3979	0.1077
Bootstrap Percentile C.I. $(\hat{\gamma}_E)$	0.881	0.3940	0.1117
Bootstrap Percentile C.I. $(\hat{\gamma}_R)$	0.885	0.3986	0.1051

Table 5.5: 90% C.I. for the scale parameter γ_0 with right censored exponential data

$X \sim \text{Exp}(1), \ Y \sim \text{Exp}(2), \ C \sim \text{Exp}(4), \ \gamma_0 = 0.5$			
Censoring % for X: 19.87%		Censoring % for Y: 33.33%	
$n_X = n_Y = 100$	Coverage	Mean Length of C.I.	s.d. Length of C.I.
Normal C.I. $(\hat{\gamma}_E)$ (3.12)	0.872	0.2821	0.0589
Normal C.I. $(\hat{\gamma}_R)$ (3.20)	0.868	0.2839	0.0571
Bootstrap Percentile C.I. $(\hat{\gamma}_E)$	0.850	0.2839	0.0571
Bootstrap Percentile C.I. $(\hat{\gamma}_R)$	0.866	0.2807	0.0567

Table 5.6: 90% C.I. for the scale parameter γ_0 with right censored exponential data

$X \sim \mathrm{Exp}(2),\ Y \sim \mathrm{Exp}(5),\ C \sim \mathrm{Exp}(8),\ \gamma_0 = 0.4$			
Censoring % for X: 20.07%		Censoring % for Y: 38.80%	
$n_X = n_Y = 25$	Coverage	Mean Length of C.I.	s.d. Length of C.I.
Normal C.I. ($\hat{\gamma}_E$) (3.12)	0.896	0.4812	0.1907
Normal C.I. ($\hat{\gamma}_R$) (3.20)	0.826	0.4469	0.1753
Bootstrap Percentile C.I. ($\hat{\gamma}_E$)	0.853	0.4642	0.1831
Bootstrap Percentile C.I. ($\hat{\gamma}_R$)	0.836	0.4276	0.1695

Table 5.7: 90% C.I. for the scale parameter γ_0 with right censored exponential data

$X \sim \mathrm{Exp}(2),\ Y \sim \mathrm{Exp}(5),\ C \sim \mathrm{Exp}(8),\ \gamma_0 = 0.4$			
Censoring % for X: 19.97%		Censoring % for Y: 38.25%	
$n_X = n_Y = 50$	Coverage	Mean Length of C.I.	s.d. Length of C.I.
Normal C.I. ($\hat{\gamma}_E$) (3.12)	0.904	0.3308	0.0922
Normal C.I. ($\hat{\gamma}_R$) (3.20)	0.883	0.3231	0.0872
Bootstrap Percentile C.I. ($\hat{\gamma}_E$)	0.874	0.3248	0.0907
Bootstrap Percentile C.I. ($\hat{\gamma}_R$)	0.886	0.3161	0.0850

Table 5.8: 90% C.I. for the scale parameter γ_0 with right censored exponential data

$X \sim \text{Exp}(2), \ Y \sim \text{Exp}(5), \ C \sim \text{Exp}(8), \ \gamma_0 = 0.4$			
Censoring % for X: 19.87%		Censoring % for Y: 38.28%	
$n_X = n_Y = 100$	Coverage	Mean Length of C.I.	s.d. Length of C.I.
Normal C.I. ($\hat{\gamma}_E$) (3.12)	0.868	0.2328	0.0494
Normal C.I. ($\hat{\gamma}_R$) (3.20)	0.863	0.2307	0.0465
Bootstrap Percentile C.I. ($\hat{\gamma}_E$)	0.846	0.2301	0.0489
Bootstrap Percentile C.I. ($\hat{\gamma}_R$)	0.865	0.2281	0.0462

The Normal confidence interval (3.12) based on $\hat{\gamma}_E$ has the best coverage level. We notice that the mean length of the Normal confidence intervals (3.12) based on $\hat{\gamma}_E$ are longer than the others for sample sizes $n = 25$ and $n = 50$. Future research can be done to study this further.

5.3   Point and Interval Estimators of the Treatment Mean

In this section, we provide simulation results on point estimates and interval estimates for the treatment mean μ_X. While only a few selected results are included in this dissertation, additional simulations yielded similar results. As in the previous section, we denote $\text{Exp}(\mu)$ as the exponential distribution with mean μ.

Table 5.9: Point Estimators for the treatment mean μ_X

Sample Size	Censoring %	Method	Estimate	SE	Error	Relative Error
$X \sim \text{Exp}(1),\ Y \sim \text{Exp}(2),\ C \sim \text{Exp}(4),\ \mu_X = 1$						
$n_X = n_Y = 25$	X: 20.1	$\hat{\mu}_E$	0.9876	0.2246	-0.0124	-0.0124
	Y: 33.4	$\hat{\mu}_R$	0.9319	0.1625	-0.0292	-0.0292
$n_X = n_Y = 50$	X: 20.0	$\hat{\mu}_E$	0.9919	0.1587	-0.0081	-0.0081
	Y: 33.3	$\hat{\mu}_R$	0.9608	0.1592	-0.0392	-0.0392
$n_X = n_Y = 100$	X: 20.0	$\hat{\mu}_E$	0.9952	0.1141	-0.0048	-0.0048
	Y: 33.3	$\hat{\mu}_R$	0.9766	0.1158	-0.0234	-0.0234

5.3.1 Point Estimators for the Treatment Mean

For this section, we present simulation results for the point estimate $\hat{\mu}_E$ in (4.3) and the point estimate $\hat{\mu}_R$ in (4.4). For $\hat{\mu}_R$, we use the results from the previous section to estimate $\hat{\gamma}_R$.

In Tables 5.9 10,000 right censored samples (1.16) of sizes 25, 50, and 100 are generated from $X \sim \text{Exp}(1)$ and $Y \sim \text{Exp}(2)$ with censoring distribution $C \sim \text{Exp}(4)$. Similarly, in Table 5.10, 10,000 right censored samples (1.16) of sizes 25, 50, and 100 are generated from $X \sim \text{Exp}(1)$ and $Y \sim \text{Exp}(2)$ with censoring distribution $C \sim \text{Exp}(4)$. For each sample size, the point estimate is given along with the censoring percentages for the X's and Y's, standard error (SE), error, and relative error.

Table 5.10: Point Estimators for the treatment mean μ_X

Sample Size	Censoring %	Method	Estimate	SE	Error	Relative Error
$X \sim \text{Exp}(2), \ Y \sim \text{Exp}(5), \ C \sim \text{Exp}(8), \ \mu_X = 2$						
$n_X = n_Y = 25$	X: 20.1	$\hat{\mu}_E$	1.9751	0.4493	-0.0249	-0.0124
	Y: 38.5	$\hat{\mu}_R$	1.8517	0.4257	-0.1483	-0.0741
$n_X = n_Y = 50$	X: 20.0	$\hat{\mu}_E$	1.9838	0.3174	-0.0162	-0.0081
	Y: 38.5	$\hat{\mu}_R$	1.9123	0.3171	-0.0877	-0.0439
$n_X = n_Y = 100$	X: 20.0	$\hat{\mu}_E$	1.9903	0.2283	-0.0097	-0.0048
	Y: 38.5	$\hat{\mu}_R$	1.9456	0.2314	-0.0544	-0.0272

From Tables 5.9 and 5.10, we see that estimator $\hat{\mu}_E$ has smaller relative errors and thus provides a better estimate for the true mean μ_X. We note that while estimator $\hat{\mu}_R$ may not be the best choice, this estimator still provides results which are comparable to $\hat{\mu}_E$. Future research can be done to study this further.

5.3.2   Interval Estimators for the Treatment Mean

In this section we present simulation results for the interval estimates for μ_X. We compute two normal-based confidence intervals described in (4.12) based on $\hat{\mu}_E$ and $\hat{\mu}_R$, respectively. In addition, we compute the bootstrap percentile confidence interval for μ_X using the bootstrap estimates $\hat{\mu}_R^*$. For each of the 1,000 simulation loops, we take 400 nested bootstrap

samples and compute the bootstrap percentile confidence interval for μ_X based on $\hat{\mu}_R$ as described in (5.3). Finally, we compute the Weighted Empirical Likelihood Ratio Confidence Interval (WELRCI) for μ_0 by $(4.88)-(4.89)$.

For Tables 5.11, 5.12, and 5.13 1,000 right censored samples (1.16) of sizes 25, 50, and 100 are generated from $X \sim \text{Exp}(1)$ and $Y \sim \text{Exp}(2)$ with censoring distribution $C \sim \text{Exp}(4)$ and the interval estimates are calculated. Similarly, Tables 5.14, 5.15, and 5.16, 1,000 right censored samples (1.16) of sizes 25, 50, and 100 are taken from $X \sim \text{Exp}(2)$ and $Y \sim \text{Exp}(5)$ with censoring distribution $C \sim \text{Exp}(8)$ and the interval estimates are calculated.

Table 5.11: 90% C.I. for the treatment mean μ_0 with right censored exponential data.

$X \sim \text{Exp}(1),\ \ Y \sim \text{Exp}(2),\ \ C \sim \text{Exp}(4),\ \ \mu_0 = 1$			
Censoring % for X: 20.07%		Censoring % for Y: 33.48%	
$n_X = n_Y = 25$	Coverage	Mean Length of C.I.	s.d. Length of C.I.
WELRCI	0.838	0.6466	0.1832
Normal C.I. $(\hat{\mu}_E)$ (4.12)	0.828	0.6510	0.2476
Normal C.I. $(\hat{\mu}_E)$ (3.20)	0.797	0.6289	0.1774
Bootstrap Percentile C.I. $(\hat{\gamma}_R)$	0.787	0.6223	0.1745

From Tables $5.11-5.14$ we see that the WELRCI has the best coverage level. In Tables 5.15 and 5.16 the results are comparable. In all cases, the coverage level is insufficient. Further research can be done to examine why we obtain these results.

5.4   Simulations for the Treatment Distribution Function

In this section, we provide simulation results comparing the different estimators that we have for the treatment distribution function. We use the uniform norm to calculate the distance between $\hat{G}(x)$ and $F_X(x)$. Note that $\hat{G}(x)$ is a discrete distribution function and $F_X(x)$ is a continuous distribution function. In particular, we have:

$$\hat{G}(x) = \sum_{i=1}^{m} \hat{p}_i^X I\{W_i^X \leq x\} \qquad \text{and} \qquad F_X(x) = 1 - e^{-x/\mu_X}$$

128

Table 5.12: 90% C.I. for the treatment mean μ_0 with right censored exponential data.

$X \sim \text{Exp}(1), \ Y \sim \text{Exp}(2), \ C \sim \text{Exp}(4), \ \mu_0 = 1$			
Censoring % for X: 19.97%		Censoring % for Y: 33.33%	
$n_X = n_Y = 50$	Coverage	Mean Length of C.I.	s.d. Length of C.I.
WELRCI	0.863	0.4974	0.1066
Normal C.I. $(\hat{\mu}_E)$ (4.12)	0.862	0.4868	0.1374
Normal C.I. $(\hat{\mu}_E)$ (3.20)	0.843	0.4862	0.1016
Bootstrap Percentile C.I. $(\hat{\gamma}_R)$	0.830	0.4827	0.1006

Table 5.13: 90% C.I. for the treatment mean μ_0 with right censored exponential data

$X \sim \text{Exp}(1), \ Y \sim \text{Exp}(2), \ C \sim \text{Exp}(4), \ \mu_0 = 1$			
Censoring % for X: 19.87%		Censoring % for Y: 33.33%	
$n_X = n_Y = 100$	Coverage	Mean Length of C.I.	s.d. Length of C.I.
WELRCI	0.863	0.3670	0.0595
Normal C.I. $(\hat{\mu}_E)$ (4.12)	0.853	0.3555	0.0763
Normal C.I. $(\hat{\mu}_E)$ (3.20)	0.842	0.3618	0.0575
Bootstrap Percentile C.I. $(\hat{\gamma}_R)$	0.828	0.3599	0.0578

Table 5.14: 90% C.I. for the treatment mean μ_0 with right censored exponential data

$X \sim \text{Exp}(2), \ Y \sim \text{Exp}(5), \ C \sim \text{Exp}(8), \ \mu_0 = 2$			
Censoring % for X: 20.07%		Censoring % for Y: 38.80%	
$n_X = n_Y = 25$	Coverage	Mean Length of C.I.	s.d. Length of C.I.
WELRCI	0.835	1.2870	0.3688
Normal C.I. $(\hat{\mu}_E)$ (4.12)	0.828	1.3019	0.4952
Normal C.I. $(\hat{\mu}_E)$ (3.20)	0.794	1.2499	0.3552
Bootstrap Percentile C.I. $(\hat{\gamma}_R)$	0.774	0.9477	0.2925

Table 5.15: 90% C.I. for the treatment mean μ_0 with right censored exponential data

$X \sim \text{Exp}(2), \ Y \sim \text{Exp}(5), \ C \sim \text{Exp}(8), \ \mu_0 = 2$			
Censoring % for X: 19.97%		Censoring % for Y: 38.25%	
$n_X = n_Y = 50$	Coverage	Mean Length of C.I.	s.d. Length of C.I.
WELRCI	0.848	0.9892	0.2143
Normal C.I. $(\hat{\mu}_E)$ (4.12)	0.862	0.9735	0.2748
Normal C.I. $(\hat{\mu}_E)$ (3.20)	0.830	0.9653	0.2033
Bootstrap Percentile C.I. $(\hat{\gamma}_R)$	0.823	0.9583	0.2020

Table 5.16: 90% C.I. for the treatment mean μ_0 with right censored exponential data

$X \sim \text{Exp}(2), \ Y \sim \text{Exp}(5), \ C \sim \text{Exp}(8), \ \mu_0 = 2$			
$X \sim \text{Exp}(2), \ Y \sim \text{Exp}(5), \ C \sim \text{Exp}(8), \ \mu_0 = 2$			
Censoring % for X: 19.87%		Censoring % for Y: 38.28%	
$n_X = n_Y = 100$	Coverage	Mean Length of C.I.	s.d. Length of C.I.
WELRCI	0.845	0.7322	0.1217
Normal C.I. $(\hat{\mu}_E)$ (4.12)	0.853	0.7109	0.1525
Normal C.I. $(\hat{\mu}_E)$ (3.20)	0.831	0.7207	0.1163
Bootstrap Percentile C.I. $(\hat{\gamma}_R)$	0.828	0.7170	0.1174

To compare the two functions, we compute the values of each function at the jump points of $\hat{G}(x)$. In particular, we compute

x	$\hat{G}(x)$	$F_X(x)$
W_1^X	$\hat{G}\left(W_1^X\right)$	$F_X\left(W_1^X\right)$
W_2^X	$\hat{G}\left(W_2^X\right)$	$F_X\left(W_2^X\right)$
$\vdots$	$\vdots$	$\vdots$
$W_{m_0-1}^X$	$\hat{G}\left(W_{m_0-1}^X\right)$	$F_X\left(W_{m_0-1}^X\right)$
$W_{m_0}^X$	$\hat{G}\left(W_{m_0}^X\right)$	$F_X\left(W_{m_0}^X\right)$

Then, we have

$$d_1 = \left\|\hat{G}(W_i^X) - F_X(W_i^X)\right\| = \max_{1 \le i \le m_0} \left|\hat{G}(W_i^X) - F_X(W_i^X)\right|$$

Similarly, we use the uniform norm to calculate the distance between $\hat{F}_n(x)$ and $F_X(x)$. Note that $\hat{F}_n(x)$ is a discrete distribution function and $F_X(x)$ is a continuous distribution function. In particular, we have:

$$\hat{F}_n(x) = \sum_{i=1}^{m} \hat{w}_i I\{\hat{W}_i \leq x\} \qquad \text{and} \qquad F_X(x) = 1 - e^{-x/\mu_X}$$

To compare the two functions, we compute the values of each function at the jump points of $\hat{G}(x)$. In particular, we compute

x	$\hat{F}_n(x)$	$F_X(x)$
W_1^X	$\hat{F}_n\left(W_1^X\right)$	$F_X\left(W_1^X\right)$
W_2^X	$\hat{F}_n\left(W_2^X\right)$	$F_X\left(W_2^X\right)$
$\vdots$	$\vdots$	$\vdots$
$W_{m_0-1}^X$	$\hat{F}_n\left(W_{m_0-1}^X\right)$	$F_X\left(W_{m_0-1}^X\right)$
$W_{m_0}^X$	$\hat{F}_n\left(W_{m_0}^X\right)$	$F_X\left(W_{m_0}^X\right)$
$\hat{\gamma}W_1^Y$	$\hat{F}_n\left(\hat{\gamma}W_1^Y\right)$	$F_X\left(\hat{\gamma}W_1^Y\right)$
$\hat{\gamma}W_2^Y$	$\hat{F}_n\left(\hat{\gamma}W_2^Y\right)$	$F_X\left(\hat{\gamma}W_2^Y\right)$
$\vdots$	$\vdots$	$\vdots$
$\hat{\gamma}W_{m_1-1}^Y$	$\hat{F}_n\left(\hat{\gamma}W_{m_1-1}^Y\right)$	$F_X\left(\hat{\gamma}W_{m_1-1}^Y\right)$
$\hat{\gamma}W_{m_1}^Y$	$\hat{F}_n\left(\hat{\gamma}W_{m_1}^Y\right)$	$F_X\left(\hat{\gamma}W_{m_1}^Y\right)$

Then, we have

$$d_2 = \left\|\hat{F}_n(W_i^X) - F_X(W_i^X)\right\| = \max\{\Delta_1, \Delta_2\}$$

Table 5.17: Estimators for Treatment Distribution: $d_1 = \|\hat{G} - F_X\|$ $\quad d_2 = \|\hat{F}_n - F_X\|$

$X \sim \text{Exp}(1)$, $Y \sim \text{Exp}(2)$, $C \sim \text{Exp}(4)$				
Sample Size	Cens. %	Distance	Mean	SE
$n_X = n_Y = 25$	X: 20.07	d_1	0.1585	0.0530
	Y: 33.48	d_2	0.1481	0.0529
$n_X = n_Y = 50$	X: 19.97	d_1	0.1168	0.0386
	Y: 33.33	d_2	0.1100	0.0362
$n_X = n_Y = 100$	X: 19.87	d_1	0.0874	0.0286
	Y: 33.33	d_2	0.0961	0.0287

where

$$\Delta_1 = \max_{1 \leq i \leq m_0} \left| \hat{F}_n(W_i^X) - F_X(W_i^X) \right|$$

and

$$\Delta_2 = \max_{1 \leq i \leq m_1} \left| \hat{F}_n(\hat{\gamma} W_i^Y) - F_X(\hat{\gamma} W_i^Y) \right|.$$

Table 5.18: Estimators for Treatment Distribution: $d_1 = \|\hat{G} - F_X\|$ $d_2 = \|\hat{F}_n - F_X\|$

$X \sim \mathrm{Exp}(2), \ Y \sim \mathrm{Exp}(5), \ C \sim \mathrm{Exp}(8)$				
Sample Size	Cens. %	Distance	Mean	SE
$n_X = n_Y = 25$	X: 20.07	d_1	0.1585	0.0530
	Y: 38.80	d_2	0.1534	0.0545
$n_X = n_Y = 50$	X: 19.97	d_1	0.1168	0.0386
	Y: 38.25	d_2	0.1179	0.0361
$n_X = n_Y = 100$	X: 19.87	d_1	0.0874	0.0286
	Y: 38.28	d_2	0.1123	0.0314

From Tables 5.17 and 5.18 we see the average distance between $\hat{F}_n$ and F_X is smaller than the distance between $\hat{G}$ and F_X for some cases. In other cases, the results are comparable.

5.5   Summary of Simulation Results

From the simulation results in Section 5.2 we see that the rank-based point estimator for the scale paramater performs better than the naive estimator. The interval estimators for the scale parameter are comparable, with the more conservative Normal C.I. based on $\hat{\gamma}_E$ having a slightly better coverage level.

From the simulation results in Section 5.3 we see that naive point estimator, $\hat{\mu}_E$, for the mean, μ_0, of the treatment group performs better than the rank-based point estimator.

However, these results are comparable. For the interval estimators for μ_0, we see that the Weight Empirical Likelihood Based Confidence Interval (WELCI) performs the best in several cases and is comparable in the other cases. In all cases, the coverage level in insufficient. Further research can be done to examine why we obtain these results.

The simulations in this dissertation consider only right censored data. More investigation is necessary for other types of censored data. Simulations confirm that the rank-based estimator is superior to the naive estimator for point parameter estimation of the scale parameter. However, they do not show superiority of the rank-based estimator for interval estimates in the case of right censored data.

CHAPTER 6
CONCLUDING REMARKS

In the present dissertation we use the Weighted Empirical Likelihood approach to study the *Accelerated Life Model* Life Model for complicated types of censored data sets, such as doubly censored data, interval censored data, and partly interval censored data. In particular, we construct tests, confidence intervals, and goodness-of-fit tests for the Accelerated Life Model in a unified way for various types of censored data. The theory can be generalized to the case of less stringent assumptions. In particular, all of the results in this dissertation can possibly be repeated with assumption that $\rho_0 = \lim_{n \to \infty} n_0/n$ and $\rho_1 = \lim_{n \to \infty} n_1/n$ instead of $\rho_0 = n_0/n$ and $\rho_1 = n_1/n$.

Simulation studies provide comparison between the standard point estimation technique (naive estimator) and the rank-based estimator suggested in the dissertation. Although both types of estimators are theoretically sound, they deliver somewhat different performance in practice. Results of the simulations are summarized in Section 5.5.

LIST OF REFERENCES

[1] Bazarra, M., Sherali, H., and Shetty, C. (1993). *Nonlinear Programming: Theory and Algorithms,* ed. 2. Wiley and Sons.

[2] Bickel, P. J. and Ren, J. (1996). The m out of n bootstrap and goodness of fit tests with double censored data. *Lecture Notes in Statist.* 109, 91-112.

[3] Bickel, P. J. and Ren, J. (2001). The bootstrap in hypthesis testing. *State of the art in statistics and probability theory. festschrift for willem r. van zwet* (m. de gunst, c. klassen and a. van der vaart, eds.) IMS, Beachwood, OH.

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

[5] Cai, T. and Cheng, S. (2004). Semiparametric regression analysis for doubly censored data. *Biometrika*, 91, 277-290.

[6] Chang, M. N. and Yang, G. L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *Ann. Statist.*, 15, 1536-1547.

[7] Chen, K., Shen, J., and Ying, Z. (2005) Rank estimation in partial linear model with censored data. *Statistica Sinica*, 15, 767-779.

[8] Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, 34, 187-220.

[9] Cox, D. R. and Oakes, D. (1984). *Analysis of Survival data.* Chapman & Hall.

[10] De Gruttola, V. and Lagakos, S. (1989). Analysis of doubly-censored survival data, with applications to aids. *Biometrics*, 45, 1-11.

[11] Efron, B. (1967). The two sample problem with censored data. *Proc. Fifth Berkeley Symp. Math. Stat. Prob.*, 4, 831-853.

[12] Efron, B. and Tibshirani, R. (1993) *An Introduction to the Bootstrap* (Chapman & Hall/CRC Monographs on Statistics & Applied Probability)

[13] Feinleib, M., Kannel, W. B., and Garrison, R. J. (1975). The framingham offspring study: Design and preliminary data. *Preventitive Medicine*, 4, 518-525.

[14] Gill, R. D. (1983). Large sample behavious of the product-limit estimator on the whole line. *Ann. Statist.*, 11, 49-58.

[15] Goeneboom, P. and Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation.* Birkhauser Verlag.

[16] Groeneboom, P. (1996). Lectures on inverse problem. In *lecture notes in mathematics.* Springer-Verlag, Berlin.

[17] Gu, M. G. and Zhang, C.-H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann. Statist.*, 21, 611-624.

[18] Huang, J. (1999). Asymptotic properties of nonparametric estimation based on partly interval-censored data. *Statistica Sinica*, 9, 501-519.

[19] Huang, J. and Wellner, J (1995). Asymptotic Normality of the NPMLE of Linear Functionals for Interval Censored Data, Case 1. *Statistica Neerlandica*, 49, 153-163.

[20] Iranpour, R. and Chacon, P. (1988). *Basico stochastic processes: The mark kac lectures.* Macmillan Publishing Company.

[21] Jin, Z., Lin, D. Y., Wei, L. J., and Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrica*, 90, 341-353.

[22] Kaplan, E. L. and Meier, P. (1958). Non parametric estimation from incomplete observations. *J. Amer. Statist. Assoc.,* 52, 457-481.

[23] Kim, M. Y., De Gruttola, V. G., and Lagakos, S. W. (1993). Analyzing doubly censored data with covariates with application to aids. *Biometrics*, 49, 13-22.

[24] Komarek, A., Lesaffre, E., and Hilton, J. (2005). Accelerated Failure Time Model for Arbitrarily Censored Data with Smoothed Error Distribution. *J. Comp. Grap. Stat.,* 14, 726-745.

[25] Komarek, A. and Lesaffre, E. (2008). Bayesian accelerated failure time model with multivariate doubly interval-censored and flexible distributional assumptions, *JASA,* 103, 523-533.

[26] Leiderman, P., Babu, B., Kagia, J., Kraemer, H., and Leiderman, G. (1973).

[27] Miller, R. (1976). Least swuared regression with censored data. *Biometrika,* 63, 449-464.

[28] Miller, R. and Halpern, J. (1982). Regression with censored data. *Biometrika*, 69, 521-531.

[29] Mykland, P. A. and Ren, J. (1996). Algorithms for computing self-consistent and maximum likelihood estimators for doubly censored data. *Ann. Statist.*, 24, 1740-1764.

[30] 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.

[31] Owen, A.B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2), 237-492.

[32] Owen, A.B. (2001). *Empirical Likelihood.* Chapman & Hall/CRC.

[33] Ren, J., (2001). Weighted Empirical Likelihoo Ratio Confidence Intervals for the mean with censored data. *Ann. Inst. Statist. Math.*, **52**: 498-516.

[34] Ren, J. (2003). Goodness of fit tests with interval censored data. Scand. J. Statist., 20, 211-226.

[35] Ren, J. (2008). Smoothed weighted empirical likelihood ratio confidence intervals for quantiles. *Bernoulli*, 14, 725-748.

[36] Ren, J. (2008). Weighted empirical likelihod in some two-sample semiparametric model with various types of censored data. *Ann. Stat.*, 36, 147-166.

[37] Ren, J. and Peer, P.G. (2000). A study on effectiveness of screening mamograms. *International Journal of Epidemiology*, 29, 803-806.

[38] Serfling, R.J. (1980). *Approximation theorems of mathematical statitstics,* ed. 2. John Wiley and Sons, Inc.

[39] Shiryaev, A. *Probability, 2nd ed.* Springer-Verlag.

[40] Stute, W. and Wang, J. L. (1993). The strong law under random censorship. *Ann. Statist.*, 21, 1591-1607.

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

[42] Turbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Amer. Statist. Assoc.*, 69, 169-173.

[43] Wellner, J. A. (1982). Asymptotic optimality of the product limit estimator. Ann. Statist., 10, 595-602.

[44] Wellner, J. A. (1995). *Interval censoring case 2: alternative hypothesis. analysis of censored data.* Institute of Mathematical Statistics, Hayward, CA.

[45] Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.*, 9, 60-62.

[46] Zhou, M. (2005). Empirical likelihood analysis of the rank estimator for the censored accelerated life model. *Biometrika*, 92, 492-498.

[47] Abramovich, F. and Angelini, C. (2006) Bayesian Maximum a posterior Multiple Testing Procedure. *Sankyaa*, **68**: 436-460.

[48] Abramovich, F., Grinshtein, V., Pensky, M. (2007) On optimality of Bayesian testimation in the normal means problem. *Annals of Statistics*, **35**: 2261-2286.

[49] Ahn, J., Todd, M., and Marron, J. (2007). Distance-weighted discrimination. *Journal of the American Statistical Association*, **102**: 1267-1272.

[50] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**: 716-723.

[51] Auer, P., Burgsteiner, H., and Maass, W. (2008). A learning rule for very simple universal approximators consisting of a single layer of perceptrons. *Neural Networks*, **21**: 786-795.

[52] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, Series B, **57**: 289-300.

[53] Berger, J. and Pericchi, R. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of Statistical Planning and Inference*, **91**: 109-122.

[54] Bickel, P, and Levina, E. (2004). Some theory for Fisher's linear discriminant function, naive Bayes, and some alternatives when there are many more variables than observations. *Bernoulli*, **10**: 989-1010.

[55] Breiman, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics*, **37**: 373-384.

[56] Bullock, T., Hopkins, C, Popper, A., Fay, R. (2005). Electroreception. Springer, New York.

[57] Candes, E. and Tao, T. (2007). The Dantzig Selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, **35** : 2313-2351.

[58] Chang, W., and Vidakovic, B. (2002). Wavelet estimation of a base-line signal from repeated noisy measurements by vertical block shrinkage. *Computer Statistical Data Analysis*, **40**: 317-328.

[59] Cheng, S. and Higham, N. (1998). A Modified Cholesky Algorithm Based on a Symmetric Indefinite Factorization. *SIAM Journal of Matrix Analysis Applications*, **19**: 1097-1110.

[60] Crampton, W.G.R., Davis, J.K., Lovejoy, N.R., Pensky, M. (2008) Multivariate classification of animal communication signals: a simulation-based comparison of alternative signal processing procedures, using electric fishes. *Journal of Physiology – Paris*, **102**: 304-321.

[61] Corless, R. M. et al. (1996). On the Lambert Function. *Adv. Comput. Math.*, **5**: 329-359.

[62] Daniels, M. and Kass, R. (1999). Nonconjugate Bayesian Estimation of Covariance Matrices and Its Use in Hierarchical Models. *Journal of the American Statistical Association*, **94**: 1254.

[63] Demartines, P. and Herault, J. (1997). Curvilinear Component Analysis: A Self-Organizing Neural Network for Nonlinear Mapping of Data Sets. *IEEE Transactions on Neural Networks*, **8**: 148-154.

[64] Donoho, D.L., Johnstone, I.M. (1994) Ideal spatial adaptation via wavelet shrinkage. *Biometrica*, **81**: 425-455.

[65] Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least Angle Regression. *Annals of Statistics*, **32**: 407-499

[66] Fan, J. and Fan, Y. High-dimensional classification using features annealed independence rules. *The Annals of Statistics*, **6**: 2605-2637.

[67] Fisher, R. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, **7**: 179-188.

[68] Furnival, G. and Wilson, R. (1974). Regression by leaps and bounds. *Technometrics*, **16**: 499-511.

[69] George, E.I. and Foster, D.P. (2000). Calibration and empirical Bayes variable selection. *Biometrica*, **84**: 731-747.

[70] Gil-Pelaez, J. Note on the inversion theorem. *Biometrica*, **38**: 481-382.

[71] Gohberg, I. and Goldberg, S. (1981). *Basic Operator Theory.* Birkhauser, Boston.

[72] Gradshteyn, I.S. and Ryzhik, I.M. (2007). Table of Integrals, Series, and Products. Seventh ed., Academic Press: Amsterdam.

[73] Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*: **82**: 711-732.

[74] Hastings, W.K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, **57**: 97-109.

[75] Higham, N. (1986). Computing the polar decomposition - with applications. *SIAM Journal of Scientific Statistical Computing*, **7**: 1160-1174.

[76] Higham, N. (1988). Computing the nearest symmetric positive semidefinite matrix. *Linear Algebra and its Applications*, **103**: 103-118.

[77] Higham, N. (2002). Computing the nearest correlation matrix - a problem from finance. *IMA Journal of Numerical Analysis*, **22**: 329-343.

[78] Hinton, G. and Salakhutdinov, R. (2006). Reducing the Dimensionality of Data with Neural Networks. *Science*, **313**: 504-507.

[79] Hoerl, A. (1962). Application of ridge analysis to regression problems. *Chemical Engineering Progress*, **58**: 54-59.

[80] James, G., Radchenko, P., and Lv. J. (2009). DASSO: connections between the Dantzig selector and lasso. *Journal of the Royal Statistical Society*, **71**: 127-142.

[81] Kass, R. and Rafferty, A. (1995). Bayes Factors. *Journal of the America Statistical Association*, **98**: 438-455.

[82] Ledoit, O. and Wolf, M. (2004). Honey, I shrunk the sample covariance matrix: problems in mean-variance optimization. *Journal of Portfolio Management*, **30**: 110-120.

[83] Mahalanobis, P.C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*, **2**: 49-55.

[84] Marron, J., Todd, M., Ahn, J. (2007). Distance Weighted Discrimination.

[85] Meinshausen, N., Rocha, G., and Yu, B. (2007). A tale of three cousins: Lasso, L2Boosting, and Dantzig (discussion on Candes and Tao's Dantzig Selector paper). *Annals of Statistics*, **25**: 2372-2384.

[86] Mengersen, K.L. and Tweedie, R.L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Annals of Statistics*, **24**: 101-121.

[87] Murre, JM; Sturdy, DP (1995). The connectivity of the brain: multi-level quantitative analysis. *Biological cybernetics*, **73**: 529-45.

[88] Paris, R.B. (2002). A uniform asymptotic expansion for the incomplete gamma function. *Journal of Computation and Applied Mathematics*, **148**: 323-339.

[89] Pearson, K. (1901). On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine*, **6**: 559-572.

[90] Rao, C.R., and Rao, M,B. (1998). *Matrix Algebra and Its Applications in Statistics and Econometrics.* World Scientific, Singapore.

[91] Richardson, S. and Green, P. (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society.* **B, 59**: 731-792.

[92] Robert, C. and Casella, G. (2004). *Monte Carlo statistical methods.* Second ed., Springer-Verlag: New York.

[93] Sarkar, S. and Chen, J. (2004). A Bayesian stepwise multiple testing procedure. Technical Report, Temple University.

[94] Schfer, J. and Korbinian, S. (2005). A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics. *Statistical Applications in Genetics and Molecular Biology*, **4**: 32.

[95] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**: 461-464.

[96] Shen, H. and Huan, J. (2008). Sparse principal component analysis via regularized low rank matrix approximation. *Journal of Multivariate Analysis*, **99**: 1015-1034.

[97] Tai, Y. C., and Speed, T.P. (2006). A multivariate empirical Bayes statistic for replicated microarray time course data. *Ann. Stat.*, **34**: 2387-2412.

[98] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal Royal Statistical Society B*, **58**: 267-288.

[99] Tibshirani, R., Hastie, T., Narasimhan, B., and Chu, G. (2002). Class Prediction by Nearest Shrunken Centroids, with Applications to DNA Microarrays. *Statistical Science*, **18**: 104-117.

[100] Torokhti, A. and Friedland, S. (2009). Towards theory of generic Principal Component Analysis. *Journal of Multivariate Analysis*, **100**: 661-669.

[101] Ververidis, D. and Kotropoulos, C. (2009). Information Loss of the Mahalanobis Distance in High Dimensions: Application to Feature Selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **31**: 2275-2281.

[102] Waller, L.A., Turnbull, B.W. and Hardin, M.J. (1995). Obtaining Distribution Functions by Numerical Inversion of Characteristic Functions with Applications. *The American Statistician*, **49-4**: 346-350.