Application Of The Empirical Likelihood Method In Proportional Hazards Model

2006

Bin He
University of Central Florida

APPLICATION OF THE EMPIRICAL LIKELIHOOD METHOD IN PROPORTIONAL HAZARDS MODEL

by

Bin He
MS in Mathematics, University of Central Florida, 2004

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
2006

Thesis Adviser:
Jian-Jian Ren

Abstract

In survival analysis, proportional hazards model is the most commonly used and the Cox model is the most popular. These models are developed to facilitate statistical analysis frequently encountered in medical research or reliability studies. In analyzing real data sets, checking the validity of the model assumptions is a key component. However, the presence of complicated types of censoring such as double censoring and partly interval-censoring in survival data makes model assessment difficult, and the existing tests for goodness-of-fit do not have direct extension to these complicated types of censored data.

In this work, we use empirical likelihood (Owen, 1988) approach to construct goodness-of-fit test and provide estimates for the Cox model with various types of censored data. Specifically, the problems under consideration are the two-sample Cox model and stratified Cox model with right censored data, doubly censored data and partly interval-censored data. Related computational issues are discussed, and some simulation results are presented. The procedures developed in the work are applied to several real data sets with some discussion.

Key words: Bootstrap, confidence interval, Cox model, doubly censored data, empirical likelihood function, goodness-of-fit test, maximum likelihood, partly interval-censored data, proportional hazards model, right censored data, survival analysis.

Acknowledgments

I wish to express my most sincere thanks and heartfelt gratitude to my advisor Dr. Jian-Jian Ren for her patience, guidance, motivation and support throughout my research and each stage of my Ph.D studies. I can never get this done without her instruction and help.

I would like to thank other members of my Thesis Committee, Drs. Marianna Pensky, Gary D. Richardson, James R. Schott and Xiaogang Su, for serving on my committee. In particular, I am grateful to Dr. Xiaogang Su for providing detailed help with software installation and coding. Also, I would especially like to thank Dr. Ram Mohapatra for his kindness and help during my entire studies in Math Department at UCF.

Many thanks to the faculty in Math Department who helped and supported me in the past 4 years, and many thanks to the staff: Norma Robles, Janice Burns, Linda Perez, Alycia Kaczuwka for their helpful assistance.

I would also like to express my wholehearted thanks to my wife, Yi Zhou, for her endless love and patience during my studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Empirical likelihood (Owen, 1988) is a nonparametric method which is developed to construct interval estimates and tests for various statistical models without assuming that the data come from a known distribution family. Its applications extend to biased sampling problems and censored data problems. Studies have shown that the empirical likelihood inference is of comparable accuracy to alternative methods. In particular, it is shown that the empirical likelihood is Bartlett-correctable for smooth function models (Diciccio, Hall and Romano, 1991). For more references, see Owen (1990, 1991), Qin and Lawless (1994), Mykland (1995) among others.

In survival analysis, interest centers on a group or groups of individuals for each of whom (or which) there is a defined point event called *failure*, occurring after a length of time called the *failure time*. The statistical models in survival analysis are developed mainly for applications in medical follow-up and reliability studies. A special source of difficulty in survival data is censoring, and right censored data are commonly seen. The most widely used model in survival analysis is *proportional hazards model*, and the most popular is the Cox model due to its adaptability in data analysis. These models are developed to facilitate statistical analysis frequently encountered in medical research or reliability studies. In analyzing real data sets, checking the validity of the model assumptions is a key component. However, the presence of complicated types of censoring such as double censoring and partly interval-censoring in survival data makes model assessment difficult, and the existing tests for goodness-of-fit do not have direct extension to these complicated types of censored data.

Throughout this thesis, we let $X_1, X_2, \cdots, X_n$ be an independently and identically distributed (i.i.d.) random sample from a continuous and nonnegative distribution function F_0, but we consider the cases when such an i.i.d sample is not completely observable due to censoring. Specifically, what we have in mind for this work includes the following types of censored data:

Right Censored Sample: The observed data are $\boldsymbol{O}_i = (V_i,\ \delta_i)$, $i = 1, 2, \cdots, n$, with

$$V_i = \begin{cases} X_i & \text{if } X_i \le C_i, \quad \delta_i = 1, \\ C_i & \text{if } X_i > C_i, \quad \delta_i = 0, \end{cases} \tag{1}$$

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

Doubly Censored Sample: The observed data are $\boldsymbol{O}_i = (V_i,\ \delta_i)$, $i = 1, 2, \cdots, n$, with

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

where C_i and D_i are the right and left censoring variables, respectively, and they are independent of X_i with $P\{D_i < C_i\} = 1$. This type of censoring has been considered by Turnbull (1974), Chang and Yang (1987), Gu and Zhang (1993), Ren (1995), Mykland and Ren (1996), among others. One recent example of doubly censored data was encountered in a study of primary breast cancer (Ren and Peer, 2000).

Partly Interval-Censored Sample:

'Case 1' Partly Interval-Censored Data: The observed data are

$$\boldsymbol{O}_i = \begin{cases} X_i & \text{if } 1 \le i \le n_1, \\ (C_i,\ \delta_i) & \text{if } n_1 + 1 \le i \le n, \end{cases} \tag{3}$$

where $\delta_i = \mathbf{I}\{X_i \le C_i\}$ and C_i is independent of X_i;

General Partly Interval-Censored Data: The observed data are

$$\boldsymbol{O}_i = \begin{cases} X_i & \text{if } \ 1 \le i \le n_1, \\ (\mathbf{C},\ \boldsymbol{\delta_i}) & \text{if } \quad n_1 + 1 \le i \le n, \end{cases} \tag{4}$$

where for N potential examination times $C_1 < \cdots < C_N$, letting $C_0 = 0$ and $C_{N+1} = \infty$, we have $\mathbf{C} = (C_1, \cdots, C_N)$ and $\boldsymbol{\delta_i} = (\delta_i^{(1)}, \cdots, \delta_i^{(N+1)})$ with $\delta_i^{(j)} = 1$, if $C_{j-1} < X_i \le C_j$; 0, elsewhere. This means that for intervals $(0, C_1],\ (C_1, C_2], \cdots,\ (C_N, \infty)$, we know which one of them X_i falls into. These two types of partly interval-censoring were considered by Huang (1999), among others. In practice, the general partly interval-censored data were encountered in Framingham Heart Disease Study (Odell, Anderson and D'Agostino; 1992), and in the study on incidence of proteinuria in insulin-dependent diabetic patients (Enevoldsen et al., 1987).

In this work, we use empirical likelihood (Owen, 1988) approach to construct goodness-of-fit tests and provide estimates for the Cox model with various types of censored data. Specifically, the problems under consideration are the two-sample Cox model and stratified Cox model with right censored data, doubly censored data and partly interval-censored data. Related computational issues are discussed, and some simulation results are presented. The problems developed in the work are applied to several real data sets with some discussion.

This thesis is organized as follows: Chapter 2 gives a brief description of parametric and nonparametric likelihood methods, and gives the nonparametric likelihood functions for right censored data and doubly censored data; Chapter 3 briefly introduces bootstrap method and its applications; Chapter 4 describes the proportional hazards model, the Cox model and stratified Cox model; Chapter 5 presents a goodness-of-fit test for the two-sample Cox model from Ren and He (2005), discusses related computation issues, and includes some simulation results and applications to three real data sets; Chapter 6 presents an estimate for the baseline distribution function in stratified Cox model from

Ren, Su and He (2006), discusses related computation issues and includes some simulation results; and Chapter 7 gives some concluding remarks.

CHAPTER 2

LIKELIHOOD INFERENCES

This chapter briefly describes the parametric and nonparametric likelihood methods, presents the likelihood functions for right censored data and doubly censored data, respectively, and reviews related asymptotic results for the *nonparametric maximum likelihood estimate* (NPMLE) $\hat{F}_n$ for the underlying lifetime distribution F_0.

2.1 Introduction

As the most important concept for inference in parametric models, likelihood can be used to derive efficient estimators and construct tests. Likelihood ratio tests can in turn be used to construct confidence intervals. Even when the data are not completely observed, or distorted, or sampled with bias, likelihood methods can be used to offset or even correct for these problems. Knowledge arising from outside of the data can also be incorporated as constraints that restricts the domain of the likelihood function, or it may be in the form of a prior distribution to be multiplied by the likelihood function. However, a problem with parametric likelihood inferences is that we might not know which parametric families the data come from. Such misspecification can cause likelihood-based estimates to be inefficient. What might be worse is that the corresponding confidence intervals and tests can fail completely.

To deal with this problem, many statisticians have turned to nonparametric inferences in order to avoid specifying a parametric family for the data. In 1988, Owen (1988) proposed empirical likelihood for the univariate mean and some other statistics, extending

earlier work of Thomas and Grunkemeier (1975) who employ a nonparametric likelihood ratio idea to construct confidence intervals for the survival probabilities. Owen's work provides nonparametric maximum likelihood estimation which has a long history in survival analysis. Owen (1988) showed that the empirical likelihood ratio statistics have a limiting chi-squared distribution in certain situations, and showed how to obtain tests and confidence limits for parameters expressed as functionals $\theta(F_0)$ of an unknown distribution function F_0.

Empirical likelihood combines the reliability of the nonparametric methods with the flexibility and effectiveness of the likelihood approach. Like other nonparametric methods, empirical likelihood inference does not require us to specify a family of distribution for the data; like parametric likelihood methods, empirical likelihood makes an automatic determination of the shape of confidence regions because it straightforwardly incorporates side information expressed through constraints or prior distribution. Empirical likelihood method easily extends to biased sampling problems and censored data problems, and it has very favorable asymptotic properties.

Empirical likelihood, as described later, provides likelihood ratio statistics for parameters by profiling a nonparametric likelihood. This approach is analogous to that used for parametric models, although it is computationally more complex. Owen (1988) showed that for i.i.d. samples, the empirical likelihood approach is applicable to quite general class of parameters $\theta(F_0)$. Also, Owen (1991) extended the empirical likelihood method to linear regression problems.

2.2 Parametric Likelihood Inference

In parametric inference, we may construct hypothesis tests and confidence regions based on the parametric likelihood ratio. As follows, we outline the framework. Let $X_1, X_2, \cdots, X_n$ be a random sample from a distribution with a p.d.f $f(x \mid \theta)$, and let

$\mathbf{X} = (X_1, X_2, \cdots, X_n)$. Then the likelihood function for parameter θ is defined by

$$L(\theta \mid \mathbf{X}) = \prod_{i=1}^{n} f(X_i \mid \theta), \tag{5}$$

and $\hat{\theta}$ is the *maximum likelihood estimator* (MLE) for θ if $L(\theta \mid \mathbf{X})$ attains its maximum at $\theta = \hat{\theta}$ over the whole parameter space Θ for θ.

For hypothesis test

$$H_0 : \theta \in \Theta_0 \qquad \text{vs.} \qquad H_1 : \theta \in \Theta_0^c, \tag{6}$$

the likelihood ratio test statistic is given by

$$R(\mathbf{X};\ \theta) = \frac{\sup\limits_{\theta \in \Theta_0} L(\theta \mid \mathbf{X})}{\sup\limits_{\theta \in \Theta} L(\theta \mid \mathbf{X})} = \frac{\sup\limits_{\theta \in \Theta_0} L(\theta \mid \mathbf{X})}{L(\hat{\theta} \mid \mathbf{X})}, \tag{7}$$

where Θ_0 is the subset of parameter space under null hypothesis.

If we consider a simpler hypothesis test

$$H_0 : \theta = \theta_0 \qquad \text{vs.} \qquad H_1 : \theta \neq \theta_0, \tag{8}$$

the likelihood ratio test statistic (7) becomes:

$$R(\mathbf{X};\ \theta) = \frac{L(\theta_0 \mid \mathbf{X})}{L(\hat{\theta} \mid \mathbf{X})}. \tag{9}$$

In (8), if H_0 holds, i.e. $\theta = \theta_0$, then $R(\mathbf{X};\ \theta)$ should be close to 1 since $\hat{\theta}$ is close to θ; if H_0 does not hold, $R(\mathbf{X};\ \theta)$ should be small as θ_0 and $\hat{\theta}$ differ. Thus, we reject H_0 when $R(\mathbf{X};\ \theta) < c$ for some predetermined constant $0 \leq c \leq 1$. In practice, c is determined as

follows: Let $0 < \alpha < 1$, then we have for $R(\mathbf{X};\ \theta)$ in (9),

$$\begin{aligned}
\alpha &= P\{\text{Type I error}\} = P\{\text{reject } H_0 \mid H_0\} \\
&= P\{R(\mathbf{X};\ \theta) \le c \mid \theta = \theta_0\} = P\{R(\mathbf{X};\ \theta_0) \le c\} \\
&= P\{-2\log R(\mathbf{X};\ \theta_0)\ \ge -2\log c\} \\
&\approx P\{\chi_1^2\ \ge\ -2\log c\},
\end{aligned} \tag{10}$$

because from Wilks's theorem (Wilks, 1938), we know that $-2\log R(\mathbf{X};\ \theta_0)$ has a limiting chi-squared distribution. In practice, c is chosen via equation (10) for desired significance level α.

From above (8) – (10), we know that the acceptance region of θ_0 is

$$A(\theta_0) = \{\mathbf{X} \mid R(\mathbf{X};\ \theta) \ge c\} = \left\{ \mathbf{X} \;\middle|\; \frac{L(\theta_0 \mid \mathbf{X})}{L(\hat{\theta} \mid \mathbf{X})} \ge c \right\}. \tag{11}$$

This can be used to construct confidence interval of θ as follows: Let

$$\lambda(\theta) = \frac{L(\theta \mid \mathbf{X})}{L(\hat{\theta} \mid \mathbf{X})}, \tag{12}$$

then the confidence interval can be constructed as

$$C(\mathbf{X}) = \{\theta :\ \lambda(\theta) \ge c\}. \tag{13}$$

To see this, we assume θ_0 is the true parameter, then $R(\mathbf{X};\ \theta) = R(\mathbf{X};\ \theta_0)$ in (9), and from (10) – (13), we have

$$\begin{aligned}
P\{\theta_0 \in C(\mathbf{X})\} &= P\{\lambda(\theta_0)\ \ge\ c\} = P\{\mathbf{X} \in A(\theta_0)\} \\
&= P\{R(\mathbf{X};\ \theta_0) \ge c\} \approx 1\ -\ \alpha.
\end{aligned} \tag{14}$$

Hence, $C(\mathbf{X})$ is a $(1-\alpha)100\%$ confidence interval for θ_0.

2.3 Empirical Likelihood Inference

As mentioned in Chapter 1, we let $X_1, X_2, \cdots, X_n$ be a random sample from distribution function F_0. Now we introduce some definitions which will be used throughout this work.

Definition 2.3.1. The *empirical distribution function* of $X_1, X_2, \cdots, X_n$ is given by

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I\{X_i \leq x\}, \quad \text{for } -\infty < x < \infty. \tag{15}$$

Definition 2.3.2. The *empirical likelihood function* (Owen, 1988) is given by

$$L(F) = \prod_{i=1}^{n} \{F(X_i) - F(X_i-)\}, \tag{16}$$

where F is any distribution function.

Note that Definition 2.3.2 reflects a very literal interpretation of the notion of likelihood. The value $L(F)$ is the probability of getting exactly the observed sample values $X_1, X_2, \cdots, X_n$. One consequence is that $L(F) = 0$ if F is a continuous distribution. Thus to have a positive nonparametric likelihood, a distribution function F must place positive probability mass on every one of the observed data point X_i's. It has been shown that the empirical d.f. F_n in (15) maximizes $L(F)$ over all distribution function F. Empirical likelihood method is analogical to the parametric likelihood method, which is briefly reviewed as follows.

For a parameter θ_0 of F_0, we often can express it as $\theta_0 = T(F_0)$, where $T(\cdot)$ is a statistical functional. For hypothesis test (6), its empirical likelihood ratio test statistic is analog to (7) given by

$$R(\mathbf{X}) = \frac{\sup\limits_{T(F)\in\Theta_0} L(F)}{\sup\limits_{T(F)\in\Theta} L(F)} = \frac{\sup\limits_{T(F)\in\Theta_0} L(F)}{L(F_n)}, \tag{17}$$

where as aforementioned, F_n is the MLE of F_0 over the whole distribution function space.

If we consider a simpler hypothesis test

$$H_0 : \theta = \theta_0 \qquad \text{vs.} \qquad H_1 : \theta \neq \theta_0, \tag{18}$$

where $\theta = T(F)$ and $\theta_0 = T(F_0)$, the empirical likelihood ratio test statistic is analog to (9) given by :

$$R(\mathbf{X}) = \frac{\sup_{T(F)=\theta_0} L(F)}{L(F_n)}. \tag{19}$$

In (18), if H_0 holds, i.e. $T(F) = T(F_0) = \theta_0$, then $R(\mathbf{X})$ should be close to 1 since F_n is close to F_0, in turn, $T(F_n) \approx T(F_0) = \theta_0$; if H_0 does not hold, $R(\mathbf{X})$ should be small because F_0 and F_n differ, i.e. $\theta_0 = T(F_0)$ and $T(F_n)$ differ. Thus, we reject H_0 when $R(\mathbf{X}) < c$ for some predetermined constant $0 \leq c \leq 1$. In practice, analog to (10), c is determined as follows: Let $0 < \alpha < 1$, then, denoting R_0 as $R(\mathbf{X})$ under H_0, we have for $R(\mathbf{X})$ in (19),

$$\begin{aligned} \alpha &= P\{\text{Type I error}\} = P\{\text{reject } H_0 \mid H_0\} \\ &= P\{R(\mathbf{X}) \leq c \mid T(F) = \theta_0\} = P\{R_0 \leq c\} \\ &= P\{-2\log R_0 \geq -2\log c\} \\ &\approx P\{\chi_1^2 \geq -2\log c\}, \end{aligned} \tag{20}$$

because Owen (1988) showed that $-2\log R_0$ has a limiting chi-squared distribution under null hypothesis in certain situations. Thus, c can be chosen via equation (20) for desired significance level α.

From above (18) – (20), we know that the acceptance region of θ_0 is analog to equation (11) given by:

$$A(\theta_0) = \{\mathbf{X} \mid R(\mathbf{X}) \geq c\} = \left\{ \mathbf{X} \;\middle|\; \frac{\sup_{T(F)=\theta_0} L(F)}{L(F_n)} \geq c \right\}. \tag{21}$$

This can be used to construct confidence interval of θ as follows. Let

$$\lambda(F) = \frac{L(F)}{L(F_n)}, \tag{22}$$

then the confidence interval can be constructed analog to equation (13) given by:

$$C(\mathbf{X}) \ = \{\theta = T(F) \mid \lambda(F) \geq c\}\,. \tag{23}$$

To see this, we note that for a rather general class of statistical functionals $T(\cdot)$, we can show that

$$\theta \in C(\mathbf{X}) \qquad \text{iff} \qquad \sup_{T(F)=\theta} \lambda(F) \geq c. \tag{24}$$

Thus, if we assume $\theta_0 = T(F_0)$ is the true parameter, then $R(\mathbf{X}) = R_0$ in (19), and from (20) – (24), we have

$$\begin{aligned} P\{\theta_0 \in C(\mathbf{X})\} &= P\{\sup_{T(F)=\theta_0} \lambda(F) \geq c\} = P\{\mathbf{X} \in A(\theta_0)\} \\ &= P\{R_0 \geq c\} \approx 1 \ - \ \alpha. \end{aligned} \tag{25}$$

Hence, $C(\mathbf{X})$ is a $(1-\alpha)100\%$ confidence interval for θ_0.

In Owen (1988), he established (25) for the mean:

$$\theta_0 = T(F_0) = \int x dF_0(x). \tag{26}$$

In fact, he showed the following theorem:

Theorem 2.3.1. *Let $X_1,\ X_2, \cdots, X_n$ be independent variables with non-degenerate distribution function F_0 with $\int |x|^3 dF_0 < \infty$. For 0 < c <1, let $F_{c,n} = \{F|\lambda(F) \geq c, F \ll F_n\}$ and define $X_{u,n} = \sup \int xdF, X_{L,n} = \inf \int xdF$ with both extrema taken over $F \in F_{c,n}$. Then,*

$$\lim_{n \to \infty} P\{X_{L,n} \leq E(X) \leq X_{U,n}\} \ = \ P(\chi^2_{(1)} \leq -2 \log\ c). \tag{27}$$

Furthermore, Owen extended (27) to M-estimates and any Fréchet differentiable statistical functional $T(\cdot)$. Empirical likelihood ratio confidence intervals make weak distributional assumptions and are justified by having asymptotically correct coverage.

2.4 Likelihood Function for Right Censored Data

Censoring occurs when we are unable to observe the response variable of interest. The commonly encountered form of a censored observation is the one in which observation begins from origin time and terminates before the outcome of interest is observed. Since the incomplete nature of the observation occurs in the right tail of the time axis, such observations are said to be *right censored*. For example, in a clinical trial, a patient may move out of town or die in an auto accident before death from the disease of interest could be observed.

Now we derive the likelihood function for F_0 for the right censored data (1). Let F_0 and F_C denote the distribution functions of X_i and C_i, respectively, and let $(v_i,\ \delta_i)$ be the observed value of $(V_i,\ \delta_i), 1\ \leq i\ \leq n$. Then, we have

$P\{\text{Observe what we observed}\}$

$$
\begin{aligned}
&= P(V_1 = v_1, \delta_1 = \delta_1, V_2 = v_2, \cdots, V_n = v_n, \delta_n = \delta_n) \\
&= \prod_{i=1}^{n} P(V_i = v_i, \delta_i = \delta_i) \\
&= \prod_{\delta_i=1} P(V_i = v_i, \delta_i = 1) \prod_{\delta_i=0} P(V_i = v_i, \delta_i = 0) \\
&= \prod_{\delta_i=1} P(X_i = v_i, X_i \le C_i) \prod_{\delta_i=0} P(C_i = v_i, X_i > C_i) \\
&= \prod_{\delta_i=1} P(X_i = v_i, C_i \ge v_i) \prod_{\delta_i=0} P(C_i = v_i, X_i > v_i) \\
&= \prod_{\delta_i=1} P(C_i \ge v_i) P(X_i = v_i) \prod_{\delta_i=0} P(X_i > v_i) P(C_i = v_i) \\
&= \prod_{\delta_i=1} [1 - F_C(v_i-)][F_0(v_i) - F_0(v_i-)] \prod_{\delta_i=0} [1 - F_0(v_i)][F_C(v_i) - F_C(v_i-)] \qquad (28) \\
&= \left(\prod_{i=1}^{n} [F_0(v_i) - F_0(v_i-)]^{\delta_i} [1 - F_0(v_i)]^{1-\delta_i} \right) \left(\prod_{i=1}^{n} [F_C(v_i) - F_C(v_i-)]^{1-\delta_i} [1 - F_C(v_i-)]^{\delta_i} \right).
\end{aligned}
$$

Since the last term of (28) does not involve F_0, we know that the likelihood function for F_0 with right censored data (1) is given by

$$
L(F) = \prod_{i=1}^{n} [F(V_i) - F(V_i-)]^{\delta_i} [1 - F(V_i)]^{1-\delta_i}, \qquad (29)
$$

because $L(F)$ is proportional to the full likelihood function derived in (28). Thus, the NPMLE for F_0 is $\hat{F}_n$ which maximizes the value of the likelihood function $L(F)$ given by (29). It has been proven that the NPMLE $\hat{F}_n$ for right-censored data is the product-limit estimator derived by Kaplan and Meier (1958). It can be written as follows:

$$
\hat{F}_n(t) = 1 - \prod_{V_{(i)} \le t} \left(1 - \frac{1}{n - (i) + 1} \right)^{\delta_{(i)}} = 1 - \prod_{V_{(i)} \le t} \left(1 - \frac{\delta_{(i)}}{n - (i) + 1} \right), \qquad (30)
$$

where $0 \leq V_{(1)} \leq V_{(2)} \cdots \leq V_{(n)} < \infty$. Note if there are ties in the $V_{(i)}$'s, the uncensored $V_{(i)}$'s ($\delta_{(i)} = 1$) are ranked ahead of the censored $V_{(i)}$'s ($\delta_{(i)} = 0$).

It is shown that $\hat{F}_n$ given by (30) is asymptotically close to F_0 uniformly in almost surely sense for right censored data (Stute and Wang, 1993), and that under certain conditions, $\sqrt{n}(\hat{F}_n - F_0)$ weakly converges to a centered Gaussian process (Gill, 1983).

2.5 Likelihood Function for Doubly Censored Data

Due to sampling methods or other factors beyond experiment control, the measurements of lifetime may be censored from above and below. Doubly censored data has been encountered in important medical studies such as breast cancer research (Ren and Peer, 2000) and African infant precocity study (Leiderman et al., 1973).

Now we derive the likelihood function for F_0 for doubly censored data (2). Let F_0, F_C and F_D denote the distribution functions of X_i, C_i and D_i, respectively, and let (v_i, δ_i)

be the observed value of $(V_i,\ \delta_i),\ 1\ \leq i\ \leq n$. Then, we have

$$
\begin{aligned}
&P\{\text{Observe what we observed}\} \\
&= P(V_1 = v_1, \delta_1 = \delta_1, V_2 = v_2, \cdots, V_n = v_n, \delta_n = \delta_n) \\
&= \prod_{i=1}^{n} P(V_i = v_i, \delta_i = \delta_i) \\
&= \prod_{\delta_i=1} P(V_i = v_i, \delta_i = 1) \prod_{\delta_i=2} P(V_i = v_i, \delta_i = 2) \prod_{\delta_i=3} P(V_i = v_i, \delta_i = 3) \\
&= \prod_{\delta_i=1} P(X_i = v_i, D_i < X_i \leq C_i) \prod_{\delta_i=2} P(C_i = v_i, X_i > C_i) \prod_{\delta_i=3} P(D_i = v_i, X_i \leq D_i) \\
&= \prod_{\delta_i=1} P(X_i = v_i, D_i < v_i \leq C_i) \prod_{\delta_i=2} P(C_i = v_i, X_i > v_i) \prod_{\delta_i=3} P(D_i = v_i, X_i \leq v_i) \\
&= \prod_{\delta_i=1} P(D_i < v_i \leq C_i) P(X_i = v_i) \prod_{\delta_i=2} P(X_i > v_i) P(C_i = v_i) \prod_{\delta_i=3} P(X_i \leq v_i) P(D_i = v_i) \\
&= \prod_{\delta_i=1} P(D_i < v_i \leq C_i) [F_0(v_i) - F_0(v_i-)] \prod_{\delta_i=2} [1 - F_0(v_i)] P(C_i = v_i) \prod_{\delta_i=3} F_0(v_i) P(D_i = v_i) \\
&= \left\{ \prod_{\delta_i=1} [F_0(v_i) - F_0(v_i-)] \prod_{\delta_i=2} [1 - F_0(v_i)] \prod_{\delta_i=3} F_0(v_i) \right\} \\
&\quad \times \left\{ \prod_{\delta_i=1} P(D_i < v_i \leq C_i) \prod_{\delta_i=2} P(C_i = v_i) \prod_{\delta_i=3} P(D_i = v_i) \right\}. \qquad (31)
\end{aligned}
$$

Since the last term of (31) does not involve F_0, we know that the likelihood function for F_0 with doubly censored data (2) is given by

$$
L(F) = \prod_{\delta_i=1} [F(v_i) - F(v_i-)] \prod_{\delta_i=2} [1 - F(v_i)] \prod_{\delta_i=3} F(v_i), \qquad (32)
$$

because $L(F)$ is proportional to the full likelihood function derived in (31). Thus, the MLE for F_0 is $\hat{F}_n$ which maximizes the value of the likelihood function $L(F)$ given by (32).

For doubly censored samples, Turnbull (1974) and Chang, and Yang (1987) gave the *self-consistent estimators* (SCE) for the survival function $\bar{F}_0 = 1 - F_0$ with grouped data and ungrouped data, respectively. Mykland and Ren (1996) showed that the NPMLE $\hat{F}_n$

uniquely exists for doubly censored data, and they established sufficient and necessary conditions for an SCE to be the NPMLE $\hat{F}_n$. Moreover, they gave a simple algorithm to compute the NPMLE $\hat{F}_n$.

It is shown that $\hat{F}_n$ is asymptotically close to F_0 uniformly in almost surely sense for doubly censored data (Gu and Zhang, 1993), and that under certain conditions, $\sqrt{n}(\hat{F}_n - F_0)$ weakly converges to a centered Gaussian process (Gu and Zhang, 1993). The estimate of the covariance function of the Gaussian process was given by Ren (1995) which includes right censored data as a special case.

CHAPTER 3

BOOTSTRAP

In this chapter, we briefly describe bootstrap method and its applications.

3.1 Introduction

By studying and synthesizing a lot of resampling ideas that were around in the history, Efron (1979) established the bootstrap method for simulation based statistical analysis. The idea of the bootstrap is to generate more new datasets through resampling the original dataset. So that we still have the information of the original data and true underlying sample properties are reproduced as closely as possible and unknown model characteristics are replaced by sample estimates.

Unlike theoretical research, bootstrap is a computer-intensive method which allows to study the performance of statistical methods by applying them repeatedly to bootstrap resampling data. Its greatest advantage lies on routinely solving problems which are far too complicated for traditional statistical analysis. Even for relatively simple problems, the bootstrap is an increasingly good data analytic tool as we are now living in a world of tremendously declining computational costs.

3.2 The Bootstrap Estimate

Let $X_1, X_2, \cdots, X_n$ be a random sample from unknown distribution function F_0, and let F_n be the empirical distribution function based on the sample $\mathbf{X} = (X_1, X_2, \cdots, X_n)$. We want to estimate a parameter of interest $\theta = T(\mathbf{X};\ F_0)$ on sample $\mathbf{X}$. Having observed

$X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n$, we can estimate $\theta = T(\mathbf{X};\ F_0)$ on $\mathbf{X}$ by $\hat{\theta} = T(\mathbf{X};\ F_n)$ based on plug-in principle. A bootstrap sample $\mathbf{X}^* = (X_1^*, X_2^*, \cdots, X_n^*)$ is defined to be a random sample of size n drawn from observed sample $\mathbf{X}$ with replacement. A bootstrap replication of $\hat{\theta}$ is $\hat{\theta}^* = T(\mathbf{X}^*;\ F_n^*)$ based solely on bootstrap sample $\mathbf{X}^*$, where F_n^* is the empirical distribution function based on bootstrap sample $\mathbf{X}^*$.

3.2.1 *The Bootstrap Estimate of Standard Error*

The bootstrap estimate of standard error of $\hat{\theta}$ is a plug-in estimate. Specifically, we denote the standard error of $\hat{\theta}$ as $SE_{F_0}(\hat{\theta})$. Then the bootstrap estimate of standard error of $\hat{\theta}$ is defined by $\widehat{SE}_B$, which is computed as follows.

(a) Select B independent bootstrap samples $\mathbf{X}^{*1}, \mathbf{X}^{*2}, \cdots, \mathbf{X}^{*B}$, each consisting of n data values randomly drawn from $\mathbf{X}$ with replacement.

(b) Evaluate the bootstrap replication corresponding to each bootstrap sample,

$$\hat{\theta}^*(b) = T(\mathbf{X}^{*b}\ ;\ F_n^{*b}), \qquad b = 1, 2, \cdots, B, \tag{33}$$

where F_n^{*b} is the empirical d.f. based on bootstrap sample $\mathbf{X}^{*b},\ b = 1, 2, \cdots, B$.

(c) Estimate the standard error $SE_{F_0}(\hat{\theta})$ by the sample standard deviation of the B replications

$$\widehat{SE}_B = \left\{ \frac{\sum_{b=1}^{B} [\hat{\theta}^*(b) - \bar{\hat{\theta}}^*]^2}{(B-1)} \right\}^{1/2}, \tag{34}$$

where

$$\bar{\hat{\theta}}^* = \sum_{b=1}^{B} \hat{\theta}^*(b)/B. \tag{35}$$

The reason why the bootstrap can work is that for large enough n, F_n becomes close to F_0. The approximation in (34) converges to $SE_{F_0}(\hat{\theta})$ as $B \to \infty$ by the law of large

numbers and some mild assumptions. In practice, we might take B large enough so that errors in (34) are negligible.

To get the confidence interval estimate of statistics through bootstrap, let's assume we are interested in the parameter $\theta = T(\mathbf{X};\ F_0)$, and $\hat{\theta} = T(\mathbf{X};\ F_n)$ is its plug-in estimator.

One type of bootstrap confidence interval is *percentile confidence interval*. Suppose the bootstrap data set $\mathbf{X}^*$'s are randomly generated, and bootstrap replications $\hat{\theta}^*$ are computed. Let $\hat{G}$ be the cumulative distribution function of $\hat{\theta}^*$. The $1-2\alpha$ percentile interval is defined by the α and $1-\alpha$ percentile of $\hat{G}$: $[\hat{G}^{-1}(\alpha),\ \hat{G}^{-1}(1-\alpha)]$. Since $\hat{G}^{-1}(\alpha) = \hat{\theta}^{*(\alpha)}$ is the αth quantile of the bootstrap distribution, we can also write the percentile interval as

$$[\hat{\theta}^{*(\alpha)},\ \hat{\theta}^{*(1-\alpha)}]. \tag{36}$$

The above expression refers to the ideal bootstrap situation in which the number of bootstrap replications is infinite. In practice, we must use some finite number B of replications, therefore, the approximate $1-2\alpha$ percentile interval is

$$[\hat{\theta}_B^{*(\alpha)},\ \hat{\theta}_B^{*(1-\alpha)}]. \tag{37}$$

To proceed, we generate B independent bootstrap datasets $\mathbf{X}^{*1}, \cdots, \mathbf{X}^{B}$ and compute the bootstrap replications $\hat{\theta}^*(b),\ b = 1, \cdots, B$, then $\hat{\theta}_B^{*(\alpha)}$ would be the αth empirical quantile of the $\hat{\theta}^*(b)$ values. That is, the $(B \cdot \alpha)$th value in the ordered list of the B replications of $\hat{\theta}^*$.

3.2.2 *Bootstrap Central Limit Theorem*

Let $X_1, X_2, \cdots, X_n$ be a random sample from distribution function F_0, and let F_n be the empirical distribution function based on $\mathbf{X} = (X_1, X_2, \cdots, X_n)$. For approximating the distribution functions of statistics $\theta(\mathbf{X};\ F_0)$, since the empirical distribution function F_n is close to F_0 for large enough n, it is reasonable for us to hope that the distribution of $\hat{\theta}^*(\mathbf{X}^*;\ F_n^*)$ is weakly asymptotically close to that of $\theta(\mathbf{X};\ F_0)$. Therefore, the distribution

of the bootstrapped statistic $\hat{\theta}^*$ can be approximated by Monte Carlo simulation. This suggestive method has been validated with limit theorems for many particular θ by Efron (1979), Bickel and Freedman (1981), among others.

Giné and Zinn (1990) offered a justification of the bootstrap for functions θ of continuous functions of the empirical measures, including the Kolmogorov-Smirnov and the Cramér-von Mises statistics (in any number of dimensions). Their work is briefly described as follows.

Let (S, ℓ, P) be a probability space, and let $X_i : (S^{\mathbb{N}}, \ell^{\mathbb{N}}, P^{\mathbb{N}}) \to (S, \ell, P)$ be the coordinate functions [i.i.d.(P)]. Denote the empirical measure as

$$P_n(w) = n^{-1} \sum_{i=1}^{n} \delta_{X_i(w)}, \tag{38}$$

for $w \in S^{\mathbb{N}}$, where δ_x denotes the measure with mass 1 at x. Let $\hat{X}_{nj}^w,\ j = 1, 2, \cdots, n$, be i.i.d. $[P_n(w)]$, and denote the empirical measure based on $\{\hat{X}_{nj}^w\}_{j=1}^n$ as

$$\hat{P}_n(w) = n^{-1} \sum_{j=1}^{n} \delta_{\hat{X}_{nj}^w}. \tag{39}$$

If $\mathcal{F}$ is a class of measurable functions on (S, ℓ) such that

$$\mathbf{F} = \sup_{f \in \mathcal{F}} |f| < \infty, \tag{40}$$

for all $s \in S$, then under some measurability on $\mathcal{F}$, the conditions

$$\int \mathbf{F}^2 dP < \infty, \tag{41}$$

and

$$\sqrt{n}(P_n \ - \ P) \to \ \mathbb{G}_p \text{ weakly in } l^\infty(\mathcal{F}), \tag{42}$$

are necessary and sufficient for

$$\sqrt{n}(\hat{P}_n(w) \ - \ P_n(w)) \ \to \ \mathbb{G} \text{ weakly in } l^\infty(\mathcal{F}), w - \text{a.s.} \tag{43}$$

for a centered Gaussian process $\mathbb{G}$ independent of w. Here, $\mathbb{G}$ coincides with $\mathbb{G}_p$, the Gaussian limit in (42).

The simple version of Giné and Zinn's theorem in our notation is:

$$\sqrt{n}(F_n^* - F_n) \stackrel{w}{\Rightarrow} \mathbb{G}, \quad \text{a.s.} \tag{44}$$

provided $\sqrt{n}(F_n - F_0) \stackrel{w}{\Rightarrow} \mathbb{G}$.

3.2.3 *Bootstrap for Censored Data*

Bickel and Ren (1996) extended the central limit theorem for the bootstrapped empirical process of Giné and Zinn (1990) to censored data. Specifically, for doubly censored data (2) or right censored data (1), Bickel and Ren (1996) showed that

$$\sqrt{n}(\hat{F}_n^* - \hat{F}_n) \stackrel{w}{\Rightarrow} \mathbb{G}, \quad \text{a.s.} \tag{45}$$

provided $\sqrt{n}(\hat{F}_n - F_0) \stackrel{w}{\Rightarrow} \mathbb{G}$, where $\hat{F}_n$ is the NPMLE based on censored data $(V_i,\ \delta_i),\ 1 \leq i \leq n$, in (1) or (2), and $\hat{F}_n^*$ is the NPMLE based on the bootstrap sample $(V_i^*,\ \delta_i^*),\ 1 \leq i \leq n$.

In Huang (1999), (45) was also established for partly interval-censored data (3) - (4).

CHAPTER 4

PROPORTIONAL HAZARDS MODEL

This chapter describes the proportional hazards model, the Cox model and stratified Cox model.

4.1 Introduction

There is a long history for studying events and time in statistical research and practice which can be dated back to the 1700's. First, we would like to introduce the definitions of the survival variable, survival function and hazard function as follows.

Definition 4.1.1. A random variable T is a *survival random variable* if an observed outcome t of T lies in the interval $[0, \infty)$. The survival function is defined as

$$\bar{F}_T(t) = P\{T \ \geq \ t\} = 1 - F_T(t). \tag{46}$$

where $F_T(t)$ is the distribution function of T.

Definition 4.1.2. The *hazard function* of T is defined by

$$h_T(t) = \lim_{\triangle \to 0+} \frac{P\{ \ t \leq T \leq t + \triangle \mid T \geq t \ \}}{\triangle}. \tag{47}$$

The study of survival functions is at the heart of survival analysis, and the hazard function is the instantaneous mortality rate by its definition.

By applying the definition of conditional probability, we have

$$h_T(t) = \lim_{\triangle \to 0+} \frac{P\{\ t \le T \le t + \triangle\ \}}{\triangle \cdot P\{\ T \ge t\ \}} = \lim_{\triangle \to 0+} \frac{F_T(t+\triangle) - F_T(t)}{\triangle \cdot \bar{F_T}(t)}$$

$$= \frac{F_T'(t)}{\bar{F_T}(t)} = \frac{f_T(t)}{\bar{F_T}(t)}, \tag{48}$$

where $f_T(t)$ is the p.d.f of T. For continuous distributions, we notice that

$$h_T(t) = \frac{F_T'(t)}{\bar{F_T}(t)} = -\frac{d}{dt}\left\{\log \bar{F_T}(t)\right\}, \tag{49}$$

and $\bar{F_T}(0) = 1$. Thus,

$$\bar{F_T}(t) = \ \exp\left(\ -\int_0^t h_T(u)du\right) = \exp\ \{-H_T(t)\}, \tag{50}$$

where $H(\cdot)$ is called the *integrated hazard function*. Furthermore, we have from (48) and (50),

$$f_T(t) = h_T(t)\ \exp\ \{-H_T(t)\}. \tag{51}$$

Therefore, once we get the hazard function, we can specify both the density and survival function, and fully determine the distribution of T.

Other reasons for studying hazard function are:

(a) It has physically meanings as immediate risk given the objective survives to time t;

(b) Hazard-based models are often convenient when there is the censoring or other incomplete observations;

(c) Sometimes it is the best way to compare two models.

The hazard function has been widely used in the survival models. For a constant vector $\mathbf{Z}$ of explanatory variables, the *proportional hazards model* is expressed as follows:

$$h(t; \mathbf{z}, \boldsymbol{\beta}) = \psi(\mathbf{z}; \boldsymbol{\beta}) h_0(t). \tag{52}$$

Here, $h_0(\cdot)$ is the hazard under the standard conditions, also called *baseline hazard function*, and $h(t; \mathbf{z}, \boldsymbol{\beta})$ is a hazard function which is associated with $h_0(t)$ through covariate $\mathbf{Z}$ and parameter $\boldsymbol{\beta}$. The proportional hazards model (52) assumes that the hazard function $h(t; \mathbf{z}, \boldsymbol{\beta})$ is proportional to $h_0(t)$ in that their ratio is constant over survival time which is $\psi(\mathbf{z}; \boldsymbol{\beta})$. The function $\psi(\mathbf{z}; \boldsymbol{\beta})$ characterizes how the hazard function changes as a function of subject covariates $\mathbf{Z}$. In the Cox model, which is discussed in the next section, $\boldsymbol{\beta}$ reflects how the covariates change on the hazard function. In (52), function $\psi(\mathbf{z}, \boldsymbol{\beta})$ must be chosen such that $h(t; \mathbf{z}, \boldsymbol{\beta}) > 0$, and when $\mathbf{Z} = \mathbf{0}$, we require $\psi(\mathbf{0}; \boldsymbol{\beta}) = 1$, so that $h(t; \mathbf{0}, \boldsymbol{\beta}) = h_0(t)$.

4.2 Cox Model

One of the most popular proportional hazards model is Cox Model as described in this section. The foundation work in this area was done by Cox (1972). This work has become a platform for building the methodology of the last 30 years. The Cox model is the most important distribution-free regression model used in survival analysis.

As a special case of proportional hazards model (52), the Cox model assumes:

$$\psi(\mathbf{z};\boldsymbol{\beta}) = e^{\boldsymbol{\beta}^T\mathbf{z}}. \tag{53}$$

Thus, the Cox model is expressed as

$$h(t;\boldsymbol{\beta},\mathbf{z}) = e^{\boldsymbol{\beta}^T\mathbf{z}}h_0(t). \tag{54}$$

One appealing part of the Cox model is its interpretation as relative risk ratio. For example, when a covariate is dichotomous, like gender with $z_1 = 1$ for males and $z_0 = 0$ for females, the hazard ratio of Cox model becomes

$$\frac{e^{\beta^T z_1}h_0(t)}{e^{\beta^T z_0}h_0(t)} = e^{\beta(z_1-z_0)} = e^{\beta}. \tag{55}$$

Intuitively, hazard is a measure of imminent risk, and it is reasonable to model this effectively. The reasons for considering the Cox model (54) are that:

(a) There is a simple easily understood interpretation to the idea that the effect of treatment is to multiply the hazard by a constant factor;

(b) In some fields empirical evidence support the assumption of proportionality of hazards in distinct treatment groups;

(c) Censoring and the occurrence of several types of failure are easily to be accommodated within this formulation and in particular the technical problems of statistical inference have a simple solution when $h_0(t)$ is arbitrary.

The usual estimator $\hat{\beta}$ for β is computed by the Newton-Raphson method and is described as follows. To make our notation simpler, we consider the scalar parameter β and a single covariate Z.

Assume we have n independent observations, let $\tau_1 < \cdots < \tau_m$ be m uncensored failure times, and the remaining $n - m$ observations are right censored. Let i denote the individual failing at τ_i, and z_i is the covariate value of $i-$th individual. We can obtain the partial likelihood function $L(\beta)$, and the MLE $\hat{\beta}$ of β is a solution of the equation

$$U(\beta) = \frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^{m} \{z_i - \mathcal{A}_i(\beta)\} = 0, \tag{56}$$

where $\mathcal{A}_i(\beta) = \left(\sum_{k \in \mathcal{R}_i} z_k e^{z_k \beta} \right) / \left(\sum_{k \in \mathcal{R}_i} e^{z_k \beta} \right)$, and $\mathcal{R}_i = \{j : t_j \geq \tau_i\}$ denotes the corresponding risk sets at time τ_i. The Newton-Raphson method is usually used to solve (56) for $\hat{\beta}$.

4.3 Two-Sample Cox Model

This is one of the specific models we are interested in this work. Comparing survival distributions often occurs in biomedical study. For example, a researcher may want to compare the survival times of two or more groups of patients exposed to different treatments. A clinical oncologist may be interested in comparing the ability of two or more treatments to prolong life or maintain health. Usually the survival time would be different for different groups. Of course, we can draw the graphs of estimated survival curves, but that is only a rough way to show the difference. It does not show whether the difference

is significant or just random variations. Thus, we need to use statistical test to compare them. As follows, we describe the two sample problems for Cox model.

For a simpler form of Cox model:

$$h(t;z) = h_0(t)\ e^{z\beta_0}, \quad z = 0, 1, \tag{57}$$

where β_0 is a regression parameter, $h_0(t)$ is an arbitrary unspecified baseline hazard function, and $h(t;z)$ is the hazard function with z as the covariate, perhaps representing control and treatment groups, in which case the parameter β_0 measures the effect of treatment.

Denote $F(t;z)$ as the distribution function corresponding to $h(t;z)$. We let

$$\begin{aligned} &X_1, X_2, \cdots, X_{n_0} \text{ be a random sample from a distribution } F(t;0) \equiv G_0(t), \\ &Y_1, Y_2, \cdots, Y_{n_1} \text{ be a random sample from a distribution } F(t;1) \equiv H_0(t), \end{aligned} \tag{58}$$

where the two samples are independent and both nonnegative. From (50) we know that X_i's satisfy

$$\bar{G}_0(t) = \ \exp\left(-\int_0^t h_0(u)du \right), \tag{59}$$

while under model (57), Y_i's satisfy

$$\bar{H}_0(t) = \ \exp\left(-\int_0^t e^{\beta} h_0(u)du \right) = \left[\bar{G}_0(t)\right]^{e^{\beta}}. \tag{60}$$

Then the two-sample Cox model (57) is equivalent to

$$\bar{H}_0(t) = [\bar{G}_0(t)]^{\gamma_0}, \tag{61}$$

where $\gamma_0 = \exp(\beta_0) > 0$, and $\bar{G}_0(t) = 1 - G_0(t)$, which is a continuous survival function.

4.4 Stratified Cox Model

Up to this point, we made the proportional hazards assumptions for the Cox model, i.e., the hazard ratio comparing any two specifications of covariates is constant. We also used a proportional hazards model with a common unspecified baseline hazard function. But this may not be true for all covariates in the real world. For example, we may have data from a study in which subjects were randomized among sites. If we account for site by including it as a covariate, the model forces the baseline hazards to be proportional across study sites. This may not be justified, and if it isn't, one possible solution is to use site as a stratification variable, whereby each site would have a separate baseline hazard function. Thus, we introduce the stratified Cox model.

The stratified Cox model is a modification of the Cox proportional hazards model that allows for control by stratification of a covariate that may not satisfy the proportional hazards (PH) assumption. By using the covariate which may not satisfy the PH assumption as stratified variable, like Sites, and keeping the covariates that satisfy the PH assumptions in the model, the stratified Cox model extends the Cox model.

The general *stratified Cox model* can be described as follows:

$$h_k(t \mid \mathbf{z}) \;=\; h_{k0}(t)\exp(\mathbf{z}^T\boldsymbol{\beta}), \qquad k = 1, 2, \cdots, N, \tag{62}$$

where $h_k(t \mid \mathbf{z})$ is the conditional hazard function of r.v. X_{ki} given $\mathbf{Z}_{ki} = \mathbf{z}$, and $h_{k0}(t)$ is the baseline hazard function for the $k-$th stratum. Here, $(X_{k1}, \mathbf{Z}_{k1}), \cdots, (X_{kn_k}, \mathbf{Z}_{kn_k})$ are i.i.d. for each $k = 1, 2, \cdots, N$, and $\mathbf{Z}_{kj}$'s are i.i.d. random vectors cross strata. Here, $h_{k0}(t)$ is allowed to be different for each stratum, but the coefficients $\boldsymbol{\beta}$ are the same for each stratum. Specifically, in Chapter 6 we will consider two strata problem, i.e. k=1, 2.

The estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ obtained from usual estimator by Newton-Raphson method is mentioned in Section 4.2, but this method only works for i.i.d. non-censored data and right censored data. It does not apply to doubly censored data and partly interval-censored data. In Chapter 6, we propose a new approach to estimate $\boldsymbol{\beta}$ which is not only applicable

to non-censored data and right censored data, but also to doubly censored data and partly interval-censored data.

CHAPTER 5

TWO-SAMPLE COX MODEL

This chapter presents a goodness-of-fit test for the two-sample Cox model from Ren and He (2005), discusses related computation issues, and includes some simulation results and applications to three real data sets.

5.1 Semi-parametric Likelihood Estimation

First, we consider the two sample Cox model expressed in (61) for noncensored data, then extend our methods to censored data later.

Following the notations in Section 4.3, we, without loss of generality, let $Z_1 < Z_2 < \cdots < Z_n$ be the ordered observations of $X_1, X_2, \cdots, X_{n_0}, Y_1, Y_2, \cdots, Y_{n_1}$ in (58), where $n = n_0 + n_1$. In order to test the validity of the Cox model (61), a *semi-parametric maximum likelihood estimator* (SPMLE) $(\tilde{\gamma},\ \tilde{G})$ for $(\gamma_0,\ G_0)$ based on two samples has been derived as follows.

The likelihood function for two-sample problem (61) is given by

$$\begin{aligned} L(\gamma, G) &= \left\{\prod_{i=1}^{n_0}[G(X_i) - G(X_i-)]\right\}\left\{\prod_{j=1}^{n_1}[H(Y_j) - H(Y_j-)]\right\} \\ &= \left\{\prod_{i=1}^{n_0}[G(X_i) - G(X_i-)]\right\}\left\{\prod_{j=1}^{n_1}\gamma[\bar{G}(Y_j)]^{\gamma-1}[G(Y_j) - G(Y_j-)]\right\} \\ &= \gamma^{n_1}\prod_{i=1}^{n} p_i[1 - G(Z_i)]^{\delta_i(\gamma-1)} = \gamma^{n_1}\prod_{i=1}^{n} p_i\left(\sum_{j=i+1}^{n+1} p_j\right)^{\delta_i(\gamma-1)}, \end{aligned} \tag{63}$$

where $p_i = G(Z_i) - G(Z_i-)$, $\delta_i = \mathbf{I}\{Z_i \in \{Y_1, Y_2, \cdots, Y_{n_1}\}\}$ for $1 \le i \le n$ and $\sum_{i=1}^{n+1} p_i = 1$. If $(\tilde{\gamma},\ \tilde{G})$ is the solution of the following optimization problem:

$$\begin{cases} \max L(\gamma, \mathbf{p}) = \gamma^{n_1} \prod_{i=1}^{n} p_i \left(\sum_{j=i+1}^{n+1} p_j \right)^{\delta_i(\gamma-1)}, \\ \text{subject to: } 0 \le p_i \le 1,\ \sum_{i=1}^{n+1} p_i = 1, \end{cases} \tag{64}$$

then $\tilde{\gamma}$ and $\tilde{G}$ are the SPMLE for γ_0 and G_0, respectively.

From Ren and He (2005), the solution of $(\tilde{\gamma},\ \tilde{G})$ in (64) is derived, and is presented as follows. Let

$$G_{n_0}(x) = n_0^{-1} \sum_{i=1}^{n_0} \mathbf{I}\{X_i \le x\}, \quad H_{n_1}(x) = n_1^{-1} \sum_{i=1}^{n_1} \mathbf{I}\{Y_i \le x\},$$

$$F_n(x) = n^{-1} \sum_{i=1}^{n} \mathbf{I}\{Z_i \le x\} = \rho_0 G_{n_0}(x) + \rho_1 H_{n_1}(x), \tag{65}$$

where $\rho_0 = n_0/n$ and $\rho_1 = n_1/n$ are assumed to remain fixed. It is shown (Ren and He, 2005) that if $\tilde{\gamma} \ge 1$ is a solution of

$$0 = \psi(\gamma) \equiv \frac{n_1}{\gamma} + n_1 n \int_0^{\infty} \bar{H}_{n_1}(x-) \log \left\{ \frac{\bar{F}_n(x) + \rho_1(\gamma-1)\bar{H}_{n_1}(x-)}{\bar{F}_n(x) + \rho_1(\gamma-1)\bar{H}_{n_1}(x-) + n^{-1}} \right\} dF_n(x), \tag{66}$$

then $\tilde{G}$ is explicitly given through $(G_{n_0},\ H_{n_1})$ by

$$\begin{aligned} \tilde{\bar{G}}(t) &= \prod_{Z_i \le t} \frac{\bar{F}_n(Z_i) + \rho_1(\tilde{\gamma}-1)\bar{H}_{n_1}(Z_i-)}{\bar{F}_n(Z_i) + \rho_1(\tilde{\gamma}-1)\bar{H}_{n_1}(Z_i-) + n^{-1}} \\ &= \exp\left\{ n \int_0^t \log\left(\frac{\bar{F}_n(x) + \rho_1(\tilde{\gamma}-1)\bar{H}_{n_1}(x-)}{\bar{F}_n(x) + \rho_1(\tilde{\gamma}-1)\bar{H}_{n_1}(x-) + n^{-1}} \right) dF_n(x) \right\}. \end{aligned} \tag{67}$$

It should be noted that the expression of the last term of (67) allows ties among Z_i's.

However, it is difficult to find the solution of $\psi(\gamma) = 0$ in practice, because $\psi(\gamma)$ is not monotone. Using Taylor's expansion of log function in (66), under some regularity

conditions, it is shown that $\psi(\gamma) = (n_1/\gamma)[-\rho_0\varphi_n(\gamma) + O_p(n^{-1}\log n)]$ for $\gamma \geq 1$ and $\sqrt{n}(\tilde{\gamma} - \tilde{\gamma}_1) = o_p(1)$ for $\varphi_n(\tilde{\gamma_1}) = 0$, where

$$\varphi_n(\gamma) = \int_0^\infty \frac{-\bar{G}_{n_0}(x)dH_{n_1}(x) + \gamma\bar{H}_{n_1}(x)dG_{n_0}(x)}{\bar{F}_n(x) + \rho_1(\gamma-1)\bar{H}_{n_1}(x-)}, \quad \gamma \geq 1. \tag{68}$$

Therefore for the rest of this chapter, $\tilde{\gamma}$ is calculated as the solution of $\varphi_n(\gamma) = 0$.

The advantage of using $\varphi_n(\gamma) = 0$ instead of $\psi(\gamma) = 0$ to find $\tilde{\gamma}$ is because if

$$\delta = \varphi_n(1) = \int_0^\infty \frac{-\bar{G}_{n_0}(x)dH_{n_1}(x) + \bar{H}_{n_1}(x)dG_{n_0}(x)}{\rho_0\bar{G}_{n_0}(x) + \rho_1\bar{H}_{n_1}(x)} \leq 0, \tag{69}$$

then $\varphi_n(\gamma) = 0$ has a unique solution on interval $[1, \infty)$ because $\varphi_n(\infty) = 1/\rho_1 > 0$ and $\varphi_n(\gamma)$ is a strictly increasing function for $\gamma \geq 1$. Thus, it is much easier to calculate $\tilde{\gamma}$ through $\varphi_n(\gamma)$ in practice. If $\delta > 0$ in (69), we can just switch the positions of G_0 and H_0 in (58) - (61), then function $\varphi_n(\gamma)$ in (68) with (ρ_0, G_{n_0}) and (ρ_1, H_{n_1}) switched has a unique solution in $[1, \infty)$ for $\varphi_n(\gamma) = 0$.

On the other hand, if $\gamma_0 = 1$ in (61), we have the usual two-sample goodness-of-fit problem, for which there are various testing methods ready to be used. Therefore, here we only focus on the case $\gamma_0 \neq 1$ in (61). To see the relation between (61) and (69), we consider:

$$\varphi(\gamma) = \int_0^\infty \frac{-\bar{G}_0(x)dH_0(x) + \gamma\bar{H}_0(x)dG_0(x)}{\rho_0\bar{G}_0(x) + \rho_1\gamma\bar{H}_0(x)}, \quad \gamma > 0. \tag{70}$$

Under (61), γ_0 is the unique solution of $\varphi(\gamma) = 0$ on interval $(0, \infty)$ because $\varphi'(\gamma) > 0$ for $\gamma > 0$, and based on the strong uniform convergence of G_{n_0} and H_{n_1}, it can be shown that $\delta = \varphi(1) + o_{a.s.}(1)$. This means that if $\gamma_0 > 1$ in (61), we have $\varphi(1) < 0$, thus in (69) we have $\delta < 0$ all but finitely often with probability 1. Hence, without loss of generality, we assume (69) and $\gamma_0 > 1$ in (61) throughout this chapter.

Asymptotic Results:

To state some related asymptotic results from Ren and He (2005), we let $\zeta > 0$ be any constant inside the support of G_0 and let $\tilde{\gamma}_\zeta$ be the solution of function $\varphi_{n,\zeta}(\gamma) = 0$ for

$$\varphi_{n,\zeta}(\gamma) = \int_0^\zeta \frac{-\bar{G}_{n_0}(x)dH_{n_1}(x) + \gamma\bar{H}_{n_1}(x)dG_{n_0}(x)}{\bar{F}_n(x) + \rho_1(\gamma-1)\bar{H}_{n_1}(x-)}, \ \gamma \geq 1. \tag{71}$$

In practice, if ζ is greater than Z_n, the largest observation of two samples, then we have $\tilde{\gamma} = \tilde{\gamma}_\zeta$ and $\tilde{G} = \tilde{G}_\zeta$. Then,

Theorem 5.1.1. *Under model (61) and the strong consistency and weakly convergence of G_{n_0} and H_{n_1},*

(i) $\tilde{\gamma} \xrightarrow{a.s.} \gamma_0$ *, as* $n \to \infty$*;*

(ii) $\sqrt{n}(\tilde{\gamma}_\zeta - \gamma_0) \xrightarrow{\mathbf{D}} N(0, \sigma_\zeta^2)$*, as* $n \to \infty$*;*

(iii) $\sqrt{n}(\tilde{G}_\zeta - G_{n_0})$ *weakly converges to a centered Gaussian process on* $[0, \zeta]$*, where* $\tilde{G}_\zeta$ *is given by (67) with* $\tilde{\gamma}$ *replaced by* $\tilde{\gamma}_\zeta$*.*

Censored Data Case:

Our proofs for Theorem 5.1.1 only rely on the following asymptotic results: $\| G_{n_0} - G_0 \| \xrightarrow{a.s.} 0$, $\| H_{n_1} - H_0 \| \xrightarrow{a.s.} 0$, and $\sqrt{n_0}\,(G_{n_0} - G_0)$ and $\sqrt{n_1}\,(H_{n_1} - H_0)$ weakly converge to centered Gaussian processes, respectively. Thus, above SPMLE can be extended to the censored data as follows.

If one of the two samples or both in (58) are subject to censoring, then based on censored data, the *nonparametric maximum likelihood estimators* (NPMLE) for G_0 and H_0 can be calculated and expressed as:

$$\hat{G}(x) = \sum_{i=1}^{m_0} \hat{p}_i^X \mathbf{I}\{W_i^X \leq x\} \ \text{ and } \ \hat{H}(x) = \sum_{i=1}^{m_1} \hat{p}_i^Y \mathbf{I}\{W_i^Y \leq x\}, \tag{72}$$

respectively, where $W_1^X < W_2^X < \cdots < W_{m_0}^X$ with $\hat{p}_i^X > 0,\ 1 \le i \le m_0$ and $W_1^Y < W_2^Y < \cdots < W_{m_1}^Y$ with $\hat{p}_i^Y > 0,\ 1 \le i \le m_1$; see Kaplan and Meier (1958) for right censored data, Mykland and Ren (1996) for doubly censored data, and Huang (1999) for partly interval-censored data. As reviewed in Chapter 2, under suitable conditions, we know that $\| \hat{G} - G_0 \| \stackrel{a.s.}{\to} 0$, $\| \hat{H} - H_0 \| \stackrel{a.s.}{\to} 0$, and $\sqrt{n_0}\ (\hat{G} - G_0)$ and $\sqrt{n_1}\ (\hat{H} - H_0)$ weakly converge to centered Gaussian processes, respectively, for right censored data (Gill, 1983; Stute and Wang, 1993), doubly censored data (Gu and Zhang, 1993) and partly interval-censored data (Huang, 1999). Therefore, the asymptotic results in Theorem 5.1.1 also apply for these censored data.

For computation, we just need to calculate (66) - (69) with $(G_{n_0},\ H_{n_1})$ replaced by $(\hat{G},\ \hat{H})$ in (72), then the SPMLE for $(\gamma_0,\ G_0)$ with censored data under model (61) can be calculated accordingly denoted as $(\hat{\gamma},\ \tilde{G}_c)$.

5.2 Goodness of Fit Test

We construct the test statistic for checking the validity of model (61) based on the following idea. There are two ways to estimate G_0: one is to use the empirical d.f. G_{n_0} of the first sample, and the other is to use both samples under model assumption (61). We use the Kolmogorov-Smirnov type statistic to measure the difference between these two estimators, which gives the goodness-of-fit test statistic. Thus, once the SPMLE $(\tilde{\gamma},\ \tilde{G})$ for $(\gamma_0,\ G_0)$ based on two samples is calculated, the following Kolmogorov-Smirnov type statistic may be used as test statistic for checking the validity of model (61).

For noncensored data, if $\delta \le 0$, then the test statistic is

$$T_n = \sqrt{n} \parallel\ \tilde{G} - G_{n_0}\ \parallel . \tag{73}$$

If $\delta > 0$, the test statistic is

$$T_n = \sqrt{n} \parallel \tilde{H} - H_{n_1} \parallel . \tag{74}$$

For censored data, the SPMLE is denoted as $(\hat{\gamma},\ \tilde{G}_c)$, and $\hat{G}$ and $\hat{H}$ are given in (72). If $\delta \leq 0$, the test statistic for censored data is

$$\hat{T}_n = \sqrt{n} \parallel \tilde{G}_c - \hat{G} \parallel . \tag{75}$$

If $\delta > 0$, the test statistic for censored data is

$$\hat{T}_n = \sqrt{n} \parallel \tilde{H}_c - \hat{H} \parallel . \tag{76}$$

In order to compute the critical value or the $p-$value for test statistic T_n or $\hat{T}_n$, we suggest the following bootstrap procedure.

Bootstrap Procedure

Noncensored Data:

Let $X_1^*, X_2^*, \cdots, X_{n_0}^*$ and $Y_1^*, Y_2^*, \cdots, Y_{n_1}^*$ be bootstrap samples with replacement drawn from $X_1, X_2, \cdots, X_{n_0}$ and $Y_1, Y_2, \cdots, Y_{n_1}$, respectively, and compute $G_{n_0}^*$ and $H_{n_1}^*$ as follows:

$$G_{n_0}^*(x) = {n_0}^{-1} \sum_{i=1}^{n_0} \mathbf{I}\{X_i^* \leq x\}, \quad H_{n_1}^*(x) = {n_1}^{-1} \sum_{i=1}^{n_1} \mathbf{I}\{Y_i^* \leq x\},$$

$$F_n^*(x) = \rho_0 G_{n_0}^*(x) + \rho_1 H_{n_1}^*(x). \tag{77}$$

Following the same procedure as aforementioned, we get

$$1 - \tilde{G}^*(t) = \exp\left\{ n \int_0^t \log\left(\frac{\bar{F}_n^*(x) + \rho_1(\tilde{\gamma}^* - 1)\bar{H}_{n_1}^*(x-)}{\bar{F}_n^*(x) + \rho_1(\tilde{\gamma}^* - 1)\bar{H}_{n_1}^*(x-) + n^{-1}} \right) dF_n^*(x) \right\}, \tag{78}$$

where $\tilde{\gamma}^*$ is the solution of $\varphi_n(\gamma) = 0$ in (68) with $(G_{n_0},\ H_{n_1})$ replaced by $(G_{n_0}^*,\ H_{n_1}^*)$ in (77).

Censored Data:

The bootstrap method described above for noncensored data also works for censored data. Let $(V_1^{X^*}, \delta_1^{X^*}), \cdots, (V_{n_0}^{X^*}, \delta_{n_0}^{X^*})$ and $(V_1^{Y^*}, \delta_1^{Y^*}), \cdots, (V_{n_1}^{Y^*}, \delta_{n_1}^{Y^*})$ be bootstrap samples with replacement drawn from censored samples $(V_1^X, \delta_1^X), \cdots, (V_{n_0}^X, \delta_{n_0}^X)$ and $(V_1^Y, \delta_1^Y), \cdots, (V_{n_1}^Y, \delta_{n_1}^Y)$, respectively, and compute $\hat{G}^*$ and $\hat{H}^*$ using (72) with the bootstrap samples $(V_i^{X^*}, \delta_i^{X^*}), 1 \le i \le n_0$ and $(V_i^{Y^*}, \delta_i^{Y^*}), 1 \le i \le n_1$, respectively. Following the same procedure as aforementioned, we get

$$1 - \tilde{G}_c^*(t) = \exp\left\{ n \int_0^t \log\left(\frac{\bar{\hat{F}}_n^*(x) + \rho_1(\tilde{\gamma}^* - 1)\bar{\hat{H}}^*(x-)}{\bar{\hat{F}}_n^*(x) + \rho_1(\tilde{\gamma}^* - 1)\bar{\hat{H}}^*(x-) + n^{-1}} \right) d\hat{F}_n^*(x) \right\}, \tag{79}$$

where $\tilde{\gamma}^*$ is the solution of $\varphi_n(\gamma) = 0$ in (68) with $(G_{n_0},\ H_{n_1})$ replaced by $(\hat{G}^*,\ \hat{H}^*)$.

Compute $p-$value:

Noncensored Data:

If $\delta^* \le 0$, the critical value or the $p-$value can be estimated by the distribution of

$$T_n^* = \sqrt{n}\ \ \| \ (\tilde{G}^* - G_{n_0}^*)\ \ -\ \ (\tilde{G} - G_{n_0})\ \ \| \ . \tag{80}$$

If $\delta^* > 0$, the critical value or the $p-$value can be estimated by the distribution of

$$T_n^* = \sqrt{n}\ \ \| \ (\tilde{H}^* - H_{n_1}^*)\ \ -\ \ (\tilde{H} - H_{n_1})\ \ \| \ . \tag{81}$$

Based on the theorem of Giné and Zinn (1990) described in Section 3.2.2, the bootstrap consistency holds here.

Censored Data:

If $\delta^* \le 0$, the critical value or the $p-$value can be estimated by the distribution of

$$\hat{T}_n^* = \sqrt{n}\ \ \| \ (\tilde{G}_c^* - \hat{G}^*)\ \ -\ \ (\tilde{G}_c - \hat{G})\ \ \| \ . \tag{82}$$

If $\delta^* > 0$, the critical value or the $p-$value can be estimated by the distribution of

$$\hat{T}_n^* = \sqrt{n} \ \| \ (\tilde{H}_c^* - \hat{H}^*) \ - \ (\tilde{H}_c - \hat{H}) \ \| \ . \tag{83}$$

Based on the theorems of Bickel and Ren (1996) and Huang (1999), the bootstrap consistency also holds here.

We note that when model assumption (61) does not hold, a minor modification of the proofs for Theorem 5.1.1 shows that: $T_n \xrightarrow{P} \infty$, as $n \to \infty$, but $\sqrt{n} \ \| \ (\tilde{G}^* - G_{n_0}^*) \ - \ (\tilde{G} - G_{n_0}) \ \|$ is still asymptotically centered Gaussian. Hence, the power of our proposed test is very good, which has been shown later in our simulation studies and analysis of real datasets.

5.3 Computation Issues

This section discusses the detailed computation procedures to calculate the test statistic.

First, calculate the NPMLE of two samples.

Case 1: *Noncensored Data*

Let $X_1, X_2, \cdots, X_{n_0}$ be a random sample from $G_0(x)$, and $Y_1, Y_2, \cdots, Y_{n_1}$ a random sample from $H_0(x)$, where Y_i's are independent from X_i's. Compute the empirical d.f.'s G_{n_0} and H_{n_1}, respectively, as in (65).

Case 2: *Right Censored Data*

Let $(V_1^X, \delta_1^X) \ , \cdots, (V_{n_0}^X, \delta_{n_0}^X)$ and $(V_1^Y, \delta_1^Y) \ , \cdots, (V_{n_1}^Y, \delta_{n_1}^Y)$ be right censored data (1) for the first and second samples in (58), respectively. Compute $\hat{G}(x)$ and $\hat{H}(x)$ using (30).

Let $W_1^X < W_2^X < \cdots < W_{m_0}^X$ be distinct values of $V_1^X, \cdots, V_{n_0}^X$. By rearranging the data, we get

$$\hat{G}(x) = \sum_{i=1}^{m_0} \hat{p}_i^X \ \mathrm{I}\{W_i^X \le x\}, \ i = 1, 2, \cdots, m_0. \tag{84}$$

with all $\hat{p}_i^X > 0$, in which $\hat{p}_1^X = \hat{G}(W_1^X)$, $\hat{p}_i^X = \hat{G}(W_i^X) - \hat{G}(W_{i-1}^X)$, for $i = 2, 3, \cdots, m_0$. $\hat{H}$ is calculated similarly using sample (V_i^Y, δ_i^Y), $1 \le i \le n_1$.

Case 3: *Doubly Censored Data*

Let $(V_1^X, \delta_1^X), \cdots, (V_{n_0}^X, \delta_{n_0}^X)$ and $(V_1^Y, \delta_1^Y), \cdots, (V_{n_1}^Y, \delta_{n_1}^Y)$ be doubly censored data (2) for the first and second samples in (58), respectively. Using the algorithm proposed by Maykland and Ren (1996), compute

$$\hat{G}(x) = \sum_{i=1}^{m_0} \hat{p}_i^X \mathbf{I}\{W_i^X \le x\} \qquad \text{and} \qquad \hat{H}(x) = \sum_{i=1}^{m_1} \hat{p}_i^Y \mathbf{I}\{W_i^Y \le x\}, \tag{85}$$

where $W_1^X < \cdots < W_{m_0}^X$ and $W_1^Y < \cdots < W_{m_1}^Y$ be distinct values of $V_i^X, 1 \le i \le n_0$ and $V_i^Y, 1 \le i \le n_1$, respectively.

It should be noted that the NPMLE $\hat{G}$ and $\hat{H}$ may not be a proper distribution function. A common convention (Efron, 1967) is to adjust probability mass on the largest observation to make $\hat{G}(W_{m_0}^X) = 1$ or $\hat{H}(W_{m_1}^Y) = 1$. This kind of adjustment of NPMLE applies to all NPMLEs in this work unless otherwise mentioned.

Now we use censored data as the example to demonstrate the detailed steps for computing test statistic $\hat{T}_n$ given in (75) - (76). Note that noncensored data follow the same steps with $(\hat{G}, \hat{H})$ replaced by (G_{n_0}, H_{n_1}).

Let

- $n = n_0 + n_1$
- $\rho_0 = n_0/n$ and $\rho_1 = n_1/n$
- $Z_1 < Z_2 < \cdots < Z_m$ are all the jump points of $\hat{G}$ and $\hat{H}$
- $\hat{F}_n(x) = \rho_0 \hat{G}(x) + \rho_1 \hat{H}(x)$

To test the hypothesis (61):

$$\mathrm{H}_0: \quad \bar{H}_0(x) \;=\; [\bar{G}_0(x)]^{\gamma_0} \qquad \text{vs.} \qquad H_1: \quad \mathrm{H}_0 \;\text{ not true,}$$

where $\gamma_0 > 0$, the test statistic is calcualted by the following steps:

Step 1: Compute

$$\delta = \int_0^\infty \frac{\bar{\hat{G}}(x)d\bar{\hat{H}}(x) - \bar{\hat{H}}(x)d\bar{\hat{G}}(x)}{\rho_0\bar{\hat{G}}(x) + \rho_1\bar{\hat{H}}(x)}. \tag{86}$$

Step 2: If $\delta \leq 0$:

(a) Compute function:

$$\varphi_n(\gamma) = \int_0^\infty \frac{-\bar{\hat{G}}(x)d\hat{H}(x) + \gamma\bar{\hat{H}}(x)d\hat{G}(x)}{\bar{\hat{F}}_n(x) + \rho_1\bar{\hat{H}}(x-)(\gamma-1)}. \tag{87}$$

(b) Compute $\hat{\gamma}$:

Find a solution of $\varphi_n(\gamma) = 0$, and denote the solution as $\hat{\gamma}$. Note that φ_n is increasing on $[1, \infty)$ with $\varphi_n(1) \leq 0$ and $\varphi_n(\infty) > 0$. We use bisection method to get $\hat{\gamma}$, and the stopping rule is when $|\varphi_n(\hat{\gamma})| < 0.001$.

(c) Compute $\tilde{G}_c$: $\tilde{G}_c$ is a SPMLE which puts probability mass on each distinct observation, therefore for $j = 1, 2, \cdots, m$,

$$1 - \tilde{G}_c(Z_j) = \exp\left\{ n \int_0^{Z_j} \log\left(\frac{\bar{\hat{F}}_n(x) + \rho_1\bar{\hat{H}}(x-)(\hat{\gamma}-1)}{\bar{\hat{F}}_n(x) + \rho_1\bar{\hat{H}}(x-)(\hat{\gamma}-1) + n^{-1}} \right) d\hat{F}_n(x) \right\}, \tag{88}$$

which can be written as:

$$\tilde{G}_c(x) \;=\; \sum_{i=1}^{m} \tilde{p}_i^X \; \mathbf{I}\left\{Z_i \leq x\right\}. \tag{89}$$

(d) Compute test statistic $\hat{T}_n$:

$$\hat{T}_n = \sqrt{n} \parallel \tilde{G}_c - \hat{G} \parallel = \sqrt{n} \sup_{0 \leq x < \infty} \mid \tilde{G}_c(x) - \hat{G}(x)|. \tag{90}$$

This can be calculated by

$$\sqrt{n} \max_{1\leq i\leq m} \mid \tilde{G}_c(Z_i) - \hat{G}(Z_i)|, \tag{91}$$

where $\hat{G}(Z_j) = \sum_{i=1}^{m_0} \hat{p}_i^X \mathbf{I}\{W_i^X \leq Z_j\},\ j = 1, 2, \cdots, m.$

Step 3: If $\delta > 0$:

(a) Compute function:

$$\varphi_n(\gamma) = \int_0^\infty \frac{-\bar{\hat{H}}(x)d\hat{G}(x) + \gamma\bar{\hat{G}}(x)d\hat{H}(x)}{\bar{\hat{F}}_n(x) + \rho_0\bar{\hat{G}}(x-)(\gamma - 1)}. \tag{92}$$

(b) Compute $\hat{\xi}$:

Find a solution of $\varphi_n(\gamma) = 0$, and denote the solution as $\hat{\xi}$. Note that φ_n is increasing on [1, ∞) with $\varphi_n(1) \leq 0$ and $\varphi_n(\infty) > 0$. We use bisection method to find the solution, and the stopping rule is when $|\varphi_n(\hat{\xi})| < 0.001$.

(c) Compute $\tilde{H}_c$: $\tilde{H}_c$ is a SPMLE which puts probability mass on each distinct observation, therefore for $j = 1, 2, \cdots, m$,

$$1 - \tilde{H}_c(Z_j) = \exp\left\{ n \int_0^{Z_j} \log\left(\frac{\bar{\hat{F}}_n(x) + \rho_0\bar{\hat{G}}(x-)(\hat{\xi} - 1)}{\bar{\hat{F}}_n(x) + \rho_0\bar{\hat{G}}(x-)(\hat{\xi} - 1) + n^{-1}} \right) d\hat{F}_n(x) \right\}, \tag{93}$$

which can be written as:

$$\tilde{H}_c(x) \ = \ \sum_{i=1}^{m} \tilde{p}_i^Y \ \mathbf{I}\ \{Z_i \leq x\}. \tag{94}$$

(d) Compute test statistic $\hat{T}_n$:

$$\hat{T}_n = \sqrt{n} \parallel \ \tilde{H}_c - \hat{H} \ \parallel = \sqrt{n} \sup_{0\leq x<\infty} \mid \tilde{H}_c(x) - \hat{H}(x) \mid = \sqrt{n} \max_{1\leq i\leq m} \mid \tilde{H}_c(Z_i) - \hat{H}(Z_i) \mid, \tag{95}$$

where $\hat{H}(Z_j) = \sum_{i=1}^{m_1} \hat{p}_i^Y \ \mathrm{I}\{W_i^Y \leq Z_j\},\ j = 1, 2, \cdots, m.$

For bootstrap sample:

As follows, we state the bootstrap procedure for censored data since noncensored sample is just a special case of censored samples.

Let $(V_1^{X^*}, \delta_1^{X^*}), \cdots, (V_{n_0}^{X^*}, \delta_{n_0}^{X^*})$ and $(V_1^{Y^*}, \delta_1^{Y^*}), \cdots, (V_{n_1}^{Y^*}, \delta_{n_1}^{Y^*})$ be bootstrap samples of censored samples $(V_1^X, \delta_1^X), \cdots, (V_{n_0}^X, \delta_{n_0}^X)$ and $(V_1^Y, \delta_1^Y), \cdots, (V_{n_1}^Y, \delta_{n_1}^Y)$, respectively, and compute $\hat{G}^*$ and $\hat{H}^*$ using the similar way as aforementioned in Case 2 and Case 3 with the bootstrap samples $(V_i^{X^*}, \delta_i^{X^*}), 1 \le i \le n_0$ and $(V_i^{Y^*}, \delta_i^{Y^*}), 1 \le i \le n_1$, respectively.

Let $\hat{F}_n^*(x) = \rho_0 \hat{G}^*(x) + \rho_1 \hat{H}^*(x)$, and let $Z_1^* < Z_2^* < \cdots < Z_{m^*}^*$ be all the jump points of $\hat{G}^*$ and $\hat{H}^*$, then the statistic $\hat{T}_n^*$ for the bootstrap sample is calculated following the same procedure as mentioned before: Calculate

$$\delta^* = \int_0^\infty \frac{\bar{\hat{G}}^*(x) d\bar{\hat{H}}^*(x) - \bar{\hat{H}}^*(x) d\bar{\hat{G}}^*(x)}{\rho_0 \bar{\hat{G}}^*(x) + \rho_1 \bar{\hat{H}}^*(x)}. \tag{96}$$

If $\delta^* \le 0$: find the solution $\hat{\gamma}^*$ of

$$0 = \varphi_n(\gamma) = \int_0^\infty \frac{-\bar{\hat{G}}^*(x) d\hat{H}^*(x) + \gamma \bar{\hat{H}}^*(x) d\hat{G}^*(x)}{\bar{\hat{F}}_n^*(x) + \rho_1 \bar{\hat{H}}^*(x-)(\gamma - 1)}. \tag{97}$$

Then, compute

$$1 - \tilde{G}_c^*(Z_j^*) = \exp\left\{ n \int_0^{Z_j^*} \log\left(\frac{\bar{\hat{F}}_n^*(x) + \rho_1 \bar{\hat{H}}^*(x-)(\hat{\gamma}^* - 1)}{\bar{\hat{F}}_n^*(x) + \rho_1 \bar{\hat{H}}^*(x-)(\hat{\gamma}^* - 1) + n^{-1}} \right) d\hat{F}_n^*(x) \right\}, \tag{98}$$

and $\hat{T}_n^*$ is calculated by

$$\begin{aligned} \hat{T}_n^* &= \sqrt{n} \parallel (\tilde{G}_c^* - \hat{G}^*) - (\tilde{G} - \hat{G}) \parallel \\ &= \sqrt{n} \sup_{0 \le x < \infty} \mid (\tilde{G}_c^*(x) - \hat{G}^*(x)) - (\tilde{G}(x) - \hat{G}(x)) \mid \\ &= \sqrt{n} \max_{1 \le i \le m} \mid (\tilde{G}_c^*(Z_i) - \hat{G}^*(Z_i)) - (\tilde{G}(Z_i) - \hat{G}(Z_i)) \mid. \end{aligned} \tag{99}$$

If $\delta^* > 0$: calculate $\tilde{H}_c^*$ similarly by switching $(\rho_0, \hat{G}^*)$ and $(\rho_1, \hat{H}^*)$ in (97) and (98), and $\hat{T}_n^*$ is calculated by

$$\begin{aligned}\hat{T}_n^* &= \sqrt{n} \parallel (\tilde{H}_c^* - \hat{H}^*) \ - \ (\tilde{H} - \hat{H}) \parallel \\ &= \sqrt{n} \sup_{0 \le x < \infty} \mid (\tilde{H}_c^*(x) - \hat{H}^*(x)) \ - \ (\tilde{H}(x) - \hat{H}(x)) \mid \\ &= \sqrt{n} \max_{1 \le i \le m} \mid (\tilde{H}_c(Z_i) - \hat{H}^*(Z_i)) \ - \ (\tilde{H}(Z_i) - \hat{H}(Z_i)) \mid. \end{aligned} \tag{100}$$

5.4 Simulation Results

In this section, we present some simulation results.

Simulation on Estimations: Let $\text{Exp}(\mu)$ represent the exponential distribution with mean μ. In our simulation studies, we consider $G_0 = \text{Exp}(1)$ and $H_0 = \text{Exp}(0.5)$ with $\gamma_0 = 2$, and generate 20,000 samples with $n_0 = 150$ and $n_1 = 100$, respectively. The simulation average of $\tilde{\gamma}$ is 2.031 with standard deviation (s.d.) 0.282, while the uniform distance between T_n and T_n^* is 0.017. The same study is repeated for $n_0 = 100$ and $n_1 = 150$, which gives the simulation average of $\tilde{\gamma}$ as 2.026 with s.d. 0.285, and $\parallel T_n - T_n^* \parallel = 0.025$. The simulation distributions of T_n and T_n^* are shown in Figure A and Figure A, which are presented in the Appendix. All these results indicate that our proposed procedures perform very well.

Simulation on Powers: To study the power of the goodness-of-fit test, we generate 1,000 samples from $G_0 = \text{Exp}(1)$ and $H_0 = \text{Exp}(0.5) + \kappa U$ with $n_0 = 150$ and $n_1 = 100$, respectively, where U represents a uniform random variable from $(0, 1)$ and κ is a constant. For each sample, 400 bootstrap samples are used to estimate the $95th$ percentile of T_n^*, which is used as the critical value for T_n. The powers of the test with different values of κ are included in Table 5.1. The same studies for test with right censored data and doubly censored data are conducted, respectively, and the results are also included in Table 5.1. In Table 5.1, for right censored sample (1), $C_G = \text{Exp}(2)$ is the right censoring variable for

the first sample, and $C_H = \text{Exp}(1)$ is the right censoring variable for the second sample; for doubly censored sample (2), $C_G = \text{Exp}(3)$ and $D_G = (2/3)C_G - 2.5$ are the right and left censoring variable for the first sample, respectively, and $C_H = \text{Exp}(1)$ and $D_H = (2/3)C_H - 2.5$ are the right and left censoring variable for the second sample, respectively.

Table 5.1: Powers of Tests with 95% Significance Level

Samples (% of censoring with $\kappa = 0$)	κ								
	-1/2	-1/4	-1/8	-1/16	0	1/16	1/8	1/4	1/2
No censoring	0.999	0.805	0.204	0.094	0.056	0.115	0.255	0.708	0.969
Right Censoring: $C_G = \text{Exp}(2)$ (33.13%) $C_H = \text{Exp}(1)$ (33.14%)	0.958	0.368	0.093	0.082	0.075	0.119	0.193	0.465	0.768
Double Censoring: $C_G = \text{Exp}(3)$ (24.93%) $D_G = (2/3)C_G$ - 2.5 (19.01%) $C_H = \text{Exp}(1)$ (33.14%) $D_H = (2/3)C_H$ - 2.5 (1.47 %)	0.976	0.411	0.116	0.088	0.084	0.117	0.216	0.467	0.796

From Table 5.1, the powers behave as expected according to Theorem 5.1.1 when there is no censoring. For right censored data and doubly censored data, the powers also behave as they should, though the efficiency of the powers is less than that when there is no censoring. But this is expected because the samples considered here are rather heavily censored with moderate sample sizes. Though not included here, our extensive simulation studies show that when sample sizes n_0 and n_1 increase, the power under the null hypothesis (i.e., $\kappa = 0$) for censored data approaches 0.05, which is the correct theoretical power. The powers test curves are shown in Figure A given in the Appendix.

5.5 Examples

In this section, we apply the proposed goodness-of-fit test to three real datasets.

***Example* 1.** In a recent study of the age-dependent growth rate of primary breast cancer (Peer et al., 1993; Ren and Peer, 2000), the age X, at which a tumor volume was developed, was observed among 236 women through biennial mammographic screening from 1981 to 1990 in Nijmegen, The Netherlands. This dataset is doubly censored; see Ren and Gu (1997) for a brief description. Among these 236 women, $n_0 = 187$ began their screening mammograms after age of 50, while $n_1 = 49$ of them began before 50. These two samples contain 56 and 23 right censored observations, and 37 and 8 left censored observations, respectively. To study the effects of the starting age of the screening mammogram in detection of breast cancer, we fit the Cox model (57) for these two doubly censored samples, and conduct the goodness-of-fit test proposed. Our calculation yields: $\tilde{\gamma} = 29.955$, $\hat{T}_n = 0.457$ and $p-$value $= 0.606$, which is based on 10,000 bootstrap samples. Thus, we can not reject the Cox model for these two doubly censored samples.

***Example* 2.** In Cox (1972), two samples of leukemia patients are presented with estimator $\hat{\beta} = 1.65$ for β_0. Using the proposed methods, our calculation yields: $\tilde{\beta} = \log \tilde{\gamma} = 1.667$, $\hat{T}_n = 0.507$ and $p-$value $= 0.722$. For the same goodness-of-fit test, Gill-Schumacher test gives a $p-$value 0.72 for the Peto-Prentice weight function (Gill and Schumacher, 1987), and the Lin test gives 0.85 and 0.32 as the $p-$values for the same weight function and its modified version, respectively (Lin, 1991). In practice, there is always a problem of which weight function to choose. Unlike these tests, our proposed test in this work does not need to choose any weight functions in its implementation.

***Example* 3.** The Gastrointestinal Tumor Study Group (1982) reported the results of a trial that compared chemotherapy with $n_1 = 45$ patients to combined chemotherapy and radiation therapy with $n_0 = 45$ patients. These two samples are right censored with 2 and 6 right censored observations, respectively. To fit these two right censored samples with the Cox model (57), our proposed test procedure here yields: $\tilde{\gamma} = 1.001$, $\hat{T}_n = 1.698$ and $p-$value $= 0.003$. Thus, we reject the model assumption in (57) for this two-sample

dataset, which is consistent with Yang and Prentice's observation that the two estimated survival curves cross (Yang and Prentice, 2005). It should be mentioned that when using the two-sample Kolmogorov-Smirnov statistic $\sqrt{n} \parallel \hat{G} - \hat{H} \parallel$ to test $G_0 = H_0$, we obtain a $p-$value 0.0025 based on 10,000 bootstrap samples.

CHAPTER 6

STRATIFIED COX MODEL

This chapter presents an estimate for the baseline distribution function in stratified Cox model from Ren, Su and He (2006), discusses related computation issues and includes some simulation results.

6.1 Estimates and Tests

As mentioned in Chapter 4, here we specifically consider stratified Cox model with two strata:

$$h_k(t \mid z) \;=\; h_{k0}(t)\exp(z\beta),\ k = 1, 2, \tag{101}$$

where $h_k(t|z)$ is the conditional hazard function of r.v. X_{ki} given $Z_{ki} = z$, and $h_{k0}(t)$ is the baseline hazard function for the $k-$th stratum. Here, $(X_{k1}, Z_{k1}), \cdots, (X_{kn_k}, Z_{kn_k})$ are i.i.d. for each k = 1,2 and Z_{kj}'s are i.i.d. random variables cross strata. We want to construct goodness-of-fit test for the following null hypothesis test:

$$\mathrm{H}_0 : \quad h_{10}(t) = h_{20}(t). \tag{102}$$

The idea of our test is that we find SPMLE $\hat{F}_1$ and $\hat{F}_2$ for the distribution function F_1 and F_2 which have $h_{10}(t)$ and $h_{20}(t)$ as the corresponding hazard functions, respectively.

Then, the test statistic is given by

$$T_n = \sqrt{n} \parallel \hat{F}_1 - \hat{F}_2 \parallel, \quad \text{where} \quad n = n_1 + n_2. \tag{103}$$

For notation simplicity, we consider $X_1, X_2, \cdots, X_n$ are i.i.d. with d.f. G_0 and

$$(X_1, Z_1), \cdots, (X_n, Z_n) \quad \text{are} \quad \text{i.i.d.} \tag{104}$$

satisfying:

$$h(t \mid z) = h_0(t) \exp(z\beta), \tag{105}$$

where $h(t \mid z)$ is the conditional hazard function of X_i given $Z_i = z$, and $h_0(t)$ is the baseline hazard function with d.f. F_0. In (105), we assume the baseline d.f. to be $\bar{F}_0(t) = \exp(-H_0(t))$. Then, (105) gives $H(x \mid z) = H_0(x) \exp(z\beta)$, in turn, we have

$$\bar{F}_X(t \mid z) = \exp(-H_0(t) \exp(z\beta)) = [\exp(-H_0(t))]^{e^{z\beta}} = [\bar{F}_0(t)]^{e^{z\beta}}, \tag{106}$$

where $F_X(\cdot \mid z)$ is the conditional d.f. of X given $Z = z$. For data in (104), we consider $(X_1, z_1), \cdots, (X_n, z_n)$, where z_i's are the realizations of Z_i's. Then, under model assumption (105), for each X_i, (106) gives

$$\bar{F}(t \mid z_i) = [\bar{F}_0(t)]^{c_i} \;\Leftrightarrow\; f(t \mid z_i) = c_i \; f_0(t) \; [\bar{F}_0(t)]^{c_i - 1}, \tag{107}$$

where $c_i = \exp(z_i\beta)$, $F(\cdot \mid z_i)$ is the conditional d.f. of X_i given $Z_i = z_i$, $f(\cdot \mid z_i)$ is the density function of $F(\cdot \mid z_i)$, and $f_0(\cdot)$ is the density function of $F_0(\cdot)$. Then, under (107), the likelihood function of X_i given $Z_i = z_i$ is given by

$$\prod_{i=1}^{n} [F(X_i \mid z_i) - F(X_i - \mid z_i)] = \prod_{i=1}^{n} c_i [F_0(X_i) - F_0(X_i -)][\bar{F}_0(X_i)]^{c_i - 1}. \tag{108}$$

Hence, the likelihood function of F_0 is proportional to

$$L(F) = \prod_{i=1}^{n} p_i \left(\sum_{j=i+1}^{n+1} p_j \right)^{c_i - 1}, \tag{109}$$

where we assume $c_i \geq 1,\ 1 \leq i \leq n$, and assume $X_1 < X_2 < \cdots < X_n$ with $p_i = F(X_i) - F(X_i-),\ 1 \leq i \leq n$, and $0 \leq p_{n+1} \leq 1$ and $F(x) = \sum_{i=1}^{n} p_i\ \mathbf{I}\{X_i \leq x\}$. The MLE for F_0 is $\hat{F}_n$ which maximizes $L(F)$ in (109).

Now we describe the procedures to calculate $\hat{F}_n$ for right censored data, which also applies for noncensored data.

For the right censored data $(V_i, \delta_i, Z_i),\ i = 1, \cdots, n$, where (V_i, δ_i) are as (1), we denote the following:

$$Q_n^{(1)}(x,z) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}\{V_i \leq x, \delta_i = 1, Z_i \leq z\}, \tag{110}$$

$$Q_n^{(0)}(x,z) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}\{V_i \leq x, \delta_i = 0, Z_i \leq z\}, \tag{111}$$

$$Q_n(x,z) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}\{V_i \leq x, Z_i \leq z\}, \tag{112}$$

$$G_n(z) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}\{Z_i \leq z\}, \tag{113}$$

$$Q_{n,Z}^{(1)}(x) = Q_n^{(1)}(x,z)/G_n(z), \tag{114}$$

$$Q_{n,Z}^{(0)}(x) = Q_n^{(0)}(x,z)/G_n(z), \tag{115}$$

$$Q_{n,Z}(x) = Q_n(x,z)/G_n(z), \tag{116}$$

where $V_1 < V_2 < \cdots < V_n$. Then, we compute the conditional NPMLE by

$$\begin{aligned} \bar{\hat{F}}_{X \mid z}(x) &= \prod_{V_i \leq x} \left(1 - \frac{Q_{n,z}^{(1)}(V_i) - Q_{n,z}^{(1)}(V_i-)}{1 - Q_{n,z}(V_i-)} \right) \\ &= \prod_{V_i \leq x} \left(1 - \frac{Q_n^{(1)}(V_i,z) - Q_n^{(1)}(V_i-,z)}{G_n(z) - Q_n(V_i-,z)} \right), \end{aligned} \tag{117}$$

which gives

$$\bar{\hat{F}}_{X \mid Z_j}(x) = \prod_{V_i \le x} \left(1 - \frac{Q_n^{(1)}(V_i, Z_j) - Q_n^{(1)}(V_{i-1}, Z_j)}{G_n(Z_j) - Q_n(V_{i-1}, Z_j)}\right)$$
$$= \prod_{V_i \le x} \left(1 - \frac{n^{-1}\delta_i \mathbf{I}\{Z_i \le Z_j\}}{G_n(Z_j) - Q_n(V_i-, Z_j)}\right). \tag{118}$$

Let $Z_{(1)} < Z_{(2)} < \cdots < Z_{(n)}$ be sorted Z_j's. Hence, for each $Z_{(j)}$, we have by (118),

$$\hat{F}_{X \mid Z_{(j)}}(x) = \sum_{i=1}^{n} \hat{p}_{ij}\ \mathbf{I}\{V_i \le x\}, \tag{119}$$

which gives

$$\hat{G}(x, Z_{(j)}) = \hat{F}_{X \mid Z_{(j)}}(x) G_n(Z_{(j)}) = \frac{j}{n} \sum_{i=1}^{n} \hat{p}_{ij}\ \mathbf{I}\{V_i \le x\}. \tag{120}$$

Note that for any $Z_{(j)} \le z \le Z_{(j+1)}$, we have $\hat{G}(x, z) = \hat{G}(x, Z_{(j)})$. Finally, we will get

$$\hat{G}(x, z) = \sum_{j=1}^{n} \sum_{i=1}^{n} \hat{q}_{ij}\ \mathbf{I}\{V_i \le x, Z_{(j)} \le z\}, \tag{121}$$

where $\hat{p}_{i0} = 0$, $Z_{(n+1)} = \infty$, $\hat{q}_{ij} = [j\hat{p}_{ij} - (j-1)\hat{p}_{i,j-1}]/n$, and $\hat{p}_{ij}$'s are calculated for $Z_{(j)}$, $1 \le j \le n$.

Then we can calculate $\hat{F}_n(t)$ based on $\hat{G}(x, z)$ and β as follows:

$$\log \bar{\hat{F}}_n(t) = n \int_0^t \log \frac{\int\int \mathbf{I}\{x \le u\} \exp(z\beta) d\hat{G}(u, z) - n^{-1}}{\int\int \mathbf{I}\{x \le u\} \exp(z\beta) d\hat{G}(u, z)} \hat{G}(dx, \infty)$$
$$= n \sum_{k=1}^{n} \left(\sum_{l=1}^{n} \hat{q}_{kl}\right) \left\{ \log \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{I}\{V_k \le V_i\} \exp(Z_{(j)}\theta) \hat{q}_{ij} \ - \ n^{-1}}{\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{I}\{V_k \le V_i\} \exp(Z_{(j)}\theta) \hat{q}_{ij}} \right\} \mathbf{I}\{V_k \le t\}, \tag{122}$$

which can be expressed as

$$\log(\bar{\hat{F}}_n(t)) = \sum_{k=1}^{n} \hat{q}_k \ \mathbf{I}\{V_k \le t\}, \tag{123}$$

where

$$\hat{q}_k \ = \ n \ (\sum_{l=1}^{n} \ \hat{q}_{kl}) \left\{ \log \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{I}\{V_k \le V_i\} \exp(Z_{(j)}\theta)\hat{q}_{ij} \ - \ n^{-1}}{\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{I}\{V_k \le V_i\} \exp(Z_{(j)}\theta)\hat{q}_{ij}} \right\}. \tag{124}$$

Hence, $\hat{F}_n$ can be written as

$$\hat{F}_n(t) = \sum_{j=1}^{n} \hat{p}_j \ \mathbf{I}\{V_j \le t\}. \tag{125}$$

NOTE: To compute $\hat{F}_n$, we need to estimate parameter β in (105). There are three ways to do it. One is $\hat{\beta}$ from usual estimator by Newton-Raphson method for the Cox model described in Section 4.2. By our likelihood method, there are two consequent new estimators $\hat{\theta}$ and $\hat{\eta}$ for β, which are described as follows. It should be noted that $\hat{\beta}$ only applied to noncensored data or right censored data, while our estimators are applicable to these types of data as well as doubly censored data and partly interval-censored data.

(a) **Use $\hat{\boldsymbol{\theta}}$ as the solution of $\boldsymbol{\phi(\theta) = 0}$.** Let

$$\phi(\theta) = \bar{Z} - \sum_{k=1}^{n} \ \hat{b}_k \left(\frac{\sum_{j=1}^{n}\sum_{i=k}^{n} \hat{q}_{ij} Z_{(j)} \exp(Z_{(j)}\theta)}{\sum_{j=1}^{n}\sum_{i=k}^{n} \hat{q}_{ij} \exp(Z_{(j)}\theta)} \right), \tag{126}$$

in which $\hat{b}_k = \sum_{j=1}^{n} \hat{q}_{kj}$, and $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$. It is shown that if $\hat{G}(x, z)$ in (121) is a proper bivariate d.f., $\phi(\theta)$ is strictly increasing function. Thus, $\hat{\theta}$ uniquely exits. To compute $\hat{\theta}$, we use bisection algorithm to find the $\hat{\theta}$ as a solution of $\phi(\theta) = 0$, and the stopping rule is when $|\phi(\hat{\theta})| < 0.001$.

(b) **Use $\hat{\boldsymbol{\eta}}$ as the solution of $\boldsymbol{\tau_n(\eta) = 0}$**. Let

$$\tau_n(\eta) = \bar{Z} + n \sum_{k=1}^{n} \hat{q}_k \left(\sum_{i=k}^{n} \sum_{j=1}^{n} \hat{q}_{ij} Z_{(j)} \exp(Z_{(j)}\eta) \right) \log \left(\frac{\sum_{i=k}^{n} \sum_{j=1}^{n} \hat{q}_{ij} Z_{(j)} \exp(Z_{(j)}\eta) - n^{-1}}{\sum_{i=k}^{n} \sum_{j=1}^{n} \hat{q}_{ij} Z_{(j)} \exp(Z_{(j)}\eta)} \right), \tag{127}$$

in which $\hat{q}_k = \sum_{j=1}^{n} \hat{q}_{kj}$, and $\bar{Z} = \frac{1}{n} \sum_{i=1}^{n} Z_i$. To compute $\hat{\eta}$, we use Newton-Raphson method to find the $\hat{\eta}$ as a solution of $\tau_n(\eta) = 0$. Let $\hat{\eta}_0$ denote the starting value, and $\hat{\eta}_m$ denote the mth iteration value, and the stopping rule is when $|\hat{\eta}_{m+1} - \hat{\eta}_m| < 0.001$.

It is shown (Ren, Su and He, 2006) that for noncensored, right censored and doubly censored data,

$$\sqrt{n}(\hat{F}_n - F_0) \stackrel{w}{\Rightarrow} \mathbb{G}_0, \tag{128}$$

where $\hat{F}_n$ is given by (125), and $\mathbb{G}_0$ is a centered Gaussian process. Thus, under H_0 in (102), we have

$$\sqrt{n}(\hat{F}_{n_1} - \hat{F}_{n_2}) \stackrel{w}{\Rightarrow} \mathbb{G}_{12}, \tag{129}$$

where $\mathbb{G}_{12}$ is a centered Gaussian process. Hence, the test statistic for goodness-of-fit test (102) is given by

$$T_n = \sqrt{n} \parallel \hat{F}_{n_1} \; - \; \hat{F}_{n_2} \parallel . \tag{130}$$

The distribution of T_n can be estimated by that of

$$T_n^* = \sqrt{n} \parallel \hat{F}_{n_1}^* \; - \; \hat{F}_{n_2}^* \parallel, \tag{131}$$

where $\hat{F}_{n_1}^*$ and $\hat{F}_{n_2}^*$ are based on bootstrap samples. Thus, the $p-$value can be estimated by the percentiles of T_n^*.

6.2 Computation Issues

Our studies show that it is difficult to compute $\hat{\eta}$ given by (127). Thus, the followings are detailed simulation procedures to calculate the estimator $\hat{\theta}$ given by (126) and statistic $T_n = \sqrt{n} \parallel \hat{F}_n - F_0 \parallel$, where $\hat{F}_n$ is given by (125).

Step 1:

Generate n observations $(V_1, \delta_1, Z_1), \cdots, (V_n, \delta_n, Z_n)$ for $k = 1$ in (101).

- Generate one uniform observation U_1 from $U(0, 1)$;
- From U_1 get one observation Z_i from Exp(1);
- Generate second uniform observation U_2 from $U(0, 1)$;
- Let $\mu = \text{Exp}(-Z_i)$, from U_2 get one observation of X_i from Exp(μ);
- Generate third uniform observation U_3 from $U(0, 1)$;
- From U_3 get one observation C_i from Exp(2);
- Get right censored observation (V_i, δ_i):

$$V_i = \begin{cases} X_i & \text{if } X_i \leq C_i \quad \delta_i = 1 \\ C_i & \text{if } X_i > C_i \quad \delta_i = 0, \end{cases} \qquad i = 1, 2, \cdots, n; \tag{132}$$

Step 2: Compute Q_n and G_n in (110)-(116) based on sample $(V_i, \delta_i, Z_i),\ 1 \leq i \leq n$.

Sort Sample (V_i, δ_i, Z_i) to make $V_1 < V_2 < \cdots < V_n$, and let $Y_1 < Y_2 < \cdots < Y_n$ be sorted Z_i's, then calculate

$$Q_n(V_k, Y_j) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}\{V_i \le V_k, Z_i \le Y_j\}, \tag{133}$$

$$Q_n^{(1)}(V_k, Y_j) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}\{V_i \le V_k, \delta_i = 1, Z_i \le Y_j\}, \tag{134}$$

$$G_n(Y_j) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}\{Z_i \le Y_j\}, j = 1, 2, \cdots, n. \tag{135}$$

Step 3: Compute $\hat{G}$ in (121) based on sample $(V_i, \delta_i, Z_i),\ 1 \le i \le n$.

Calculate

$$\begin{aligned} \bar{\hat{F}}_{X|Y_j}(V_k) &= \prod_{V_i \le V_k} \left(1 - \frac{Q_n^{(1)}(V_i, Y_j) - Q_n^{(1)}(V_i-, Y_j)}{G_n(Y_j) - Q_n(V_i-,\ Y_j)}\right) \\ &= \prod_{V_i \le V_k} \left(1 - \frac{Q_n^{(1)}(V_i, Y_j) - Q_n^{(1)}(V_{i-1}, Y_j)}{G_n(Y_j) - Q_n(V_{i-1},\ Y_j)}\right). \end{aligned} \tag{136}$$

If $G_n(Y_j) - Q_n(V_{i-1},\ Y_j) = 0$, then let $\dfrac{Q_n^{(1)}(V_i, Y_j) - Q_n^{(1)}(V_{i-1}, Y_j)}{G_n(Y_j) - Q_n(V_{i-1},\ Y_j)} = 0$. Hence, for each Y_j, compute

$$\hat{F}_{X|Y_j}(x) = \sum_{k=1}^{n} \hat{p}_{kj}\ \mathbf{I}\{V_k \le x\},\ j = 1, 2, \cdots, n, \tag{137}$$

where $\hat{p}_{kj} = \hat{F}_{X|Y_j}(V_k) - \hat{F}_{X|Y_j}(V_{k-1})$ for $j = 2, 3, \cdots, n$, and $\hat{p}_{1j} = \hat{F}_{X|Y_j}(V_1)$.

Adjustment for $\hat{p}_{kj}$:

- For fixed Y_j, find $V_l = \max\{V_i \mid Z_i \le Y_j\}$.
- Then recalculate this $\hat{p}_{lj}$ to make $\sum_{k=1}^{l} \hat{p}_{kj} = 1$.

Calculate

$$\hat{G}(x, z) = \sum_{i=1}^{n}\sum_{j=1}^{n} \hat{q}_{ij}\ \mathbf{I}\{V_i \le x, Y_j \le z\}, \tag{138}$$

where $\hat{q}_{ij} = \dfrac{j}{n}\hat{p}_{ij} - \dfrac{j-1}{n}\hat{p}_{i(j-1)}, i = 1, 2, \cdots, n; j = 1, 2, \cdots, n;\ \hat{p}_{i0} = 0.$

Note: Some $\hat{q}_{ij}$ may be negative in the calculation. The following adjustment is made on $\hat{q}_{ij}$ so that $\hat{G}(x,z)$ in (138) is a proper bivariate d.f..

Adjustment for $\hat{q}_{ij}$:

- Let Δ= sum of all negative $\hat{q}_{ij}$'s;
- Let $\Delta_t = 1 - \Delta$ = sum of all positive $\hat{q}_{ij}$'s;
- Rewrite

$$\hat{G}(x,z) \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} \hat{q}'_{ij}\; \mathbf{I}\{V_i \le x, Y_j \le z\},$$

where

$$\hat{q}'_{ij} \;=\; \begin{cases} 0 & \text{if } \hat{q}_{ij} \le 0, \\ \hat{q}_{ij}(1 - \frac{|\Delta|}{\Delta_t}) & \text{if } \hat{q}_{ij} > 0. \end{cases} \tag{139}$$

For the rest of this section, we still use $\hat{q}_{ij}$ to represent $\hat{q}'_{ij}$.

Step 4: Compute $\hat{F}_n(x)$ based on $\hat{G}$ and $\hat{\theta}$.

First calculate $\hat{\theta}$: Let

$$\phi(\theta) = \bar{Z} - \sum_{k=1}^{n}\; \hat{b}_k \left(\frac{\sum_{j=1}^{n}\sum_{i=k}^{n} \hat{q}_{ij} Y_j \exp(Y_j\theta)}{\sum_{j=1}^{n}\sum_{i=k}^{n} \hat{q}_{ij} \exp(Y_j\theta)} \right), \tag{140}$$

in which $\hat{b}_k = \sum_{j=1}^{n} \hat{q}_{kj}$, and $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$. Note that $\phi(\theta)$ is a strictly increasing function when $\hat{G}$ is a proper bivariate d.f., therefore we can use bisection algorithm to find the $\hat{\theta}$ as a solution of $\phi(\theta) = 0$. The stopping rule used is when $|\phi(\hat{\theta})| < 0.001$.

Then calculate $\hat{q}_k$:

$$\hat{q}_k \;=\; n\,(\sum_{l=1}^{n}\; \hat{q}_{kl}) \left\{ \log \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{I}\{V_k \le V_i\} \exp(Y_j\hat{\theta})\hat{q}_{ij} \;-\; n^{-1}}{\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{I}\{V_k \le V_i\} \exp(Y_j\hat{\theta})\hat{q}_{ij}} \right\}. \tag{141}$$

If $\sum_{l=1}^{n} \hat{q}_{kl} = 0$, then let $\hat{q}_k = 0$.

If $\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{I}\{V_k \le V_i\} \exp(Y_j\hat{\theta})\hat{q}_{ij} = 0$, then let $\hat{q}_k = 0$.

Calculate $\hat{F}_n$:

$$\log(\bar{\hat{F}}_n(V_j)) = \sum_{k=1}^{n} \hat{q}_k \mathbf{I}\{V_k \le V_j\},\ j = 1, 2, \cdots, n. \tag{142}$$

Therefore,

$$\hat{F}_n(V_j) = 1\ -\ \exp\left(\sum_{k=1}^{j} \hat{q}_k\right),\ j = 1, 2, \cdots, n, \tag{143}$$

which can also be written as

$$\hat{F}_n(t) = \sum_{j=1}^{n} \hat{p}_j\ \mathbf{I}\{V_j \le t\}. \tag{144}$$

Step 5: Compute statistic $T_n = \sqrt{n} \parallel \hat{F}_n - F_0 \parallel$:

Calculate

$$F_0(V_i) = 1\ - \exp(-V_i), \quad i = 1, 2, \cdots, n. \tag{145}$$

Then, compute

$$\begin{aligned} T_n &= \sqrt{n} \parallel \hat{F}_n(V_i) - F_0(V_i) \parallel \\ &= \sqrt{n}\ \max|\ \hat{F}_n(V_i) - F_0(V_i)|, \quad i = 1, 2, \cdots, n. \end{aligned} \tag{146}$$

For bootstrap sample:

Step 6: Generate bootstrap sample $(V_i^*, \delta_i^*, Z_i^*),\ i = 1, \cdots, n$, from sample $(V_i, \delta_i, Z_i),\ 1 \le i \le n$.

Step 7: Compute Q_n^* and G_n^* based on sample $(V_i^*, \delta_i^*, Z_i^*),\ i = 1, \cdots, n$.

Let $W_1^* < W_2^* < \cdots < W_{m^*}^*$ be distinct values of V_i^*'s, and $Y_1^* < Y_2^* < \cdots < Y_{m_0^*}^*$ be distinct values of Z_i^*'s. Calculate

$$Q_n^*(W_k^*, Y_j^*) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}\{V_i^* \le W_k^*, Z_i^* \le Y_j^*\},\ k = 1, \cdots, m^*,\ j = 1, \cdots, m_0^*, \tag{147}$$

and calculate

$$Q_n^{*(1)}(W_k^*, Y_j^*) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}\{V_i^* \le W_k^*, \delta_i^* = 1, Z_i^* \le Y_j^*\},\ k = 1, \cdots, m^*,\ j = 1, \cdots, m_0^*. \tag{148}$$

Then, calculate

$$G_n^*(Y_j^*) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}\{Z_i^* \le Y_j^*\},\quad j = 1, 2, \cdots, m_0^*. \tag{149}$$

Step 8: Compute $\hat{G}^*$ based on sample $(V_i^*, \delta_i^*, Z_i^*),\ i = 1, \cdots, n$.

Calculate

$$\begin{aligned}\bar{\hat{F}}_{X|Y_j^*}^*(W_k^*) &= \prod_{W_i^* \le W_k^*} \left(1 \ - \ \frac{Q_n^{*(1)}(W_i^*, Y_j^*) - Q_n^{*(1)}(W_i^*-, Y_j^*)}{G_n^*(Y_j^*) \ - \ Q_n^*(W_i^*-,\ Y_j^*)}\right) \\ &= \prod_{W_i^* \le W_k^*} \left(1 \ - \ \frac{Q_n^{*(1)}(W_i^*, Y_j^*) - Q_n^{*(1)}(W_{i-1}^*, Y_j^*)}{G_n^*(Y_j^*) \ - \ Q_n^*(W_{i-1}^*,\ Y_j^*)}\right).\end{aligned} \tag{150}$$

where $k = 1, 2, \cdots, m^*,\ j = 1, 2, \cdots, m_0^*$. If $G_n^*(Y_j^*) \ - \ Q_n^*(W_{i-1}^*,\ Y_j^*) \ = \ 0$, then let $\dfrac{Q_n^{*(1)}(W_i^*, Y_j^*) - Q_n^{*(1)}(W_{i-1}^*, Y_j^*)}{G_n^*(Y_j^*) \ - \ Q_n^*(W_{i-1}^*,\ Y_j^*)} = 0$. Thus, for each distinct Y_j^* we have

$$\hat{F}_{X|Y_j^*}^*(x) = \sum_{i=1}^{m^*} \hat{p}_{ij}^*\ \mathbf{I}\{W_i^* \le x\},\ j = 1, 2, \cdots, m_0^*, \tag{151}$$

where $\hat{p}_{kj}^* = \hat{F}_{X|Y_j^*}^*(V_k^*) - \hat{F}_{X|Y_j^*}^*(V_{k-1}^*)$ for $\ j = 2, 3, \cdots, m_0^*$, and $\hat{p}_{1j}^* = \hat{F}_{X|Y_j^*}^*(V_1^*)$.

In turn, for each distinct W_k^* given Y_j^* we have

$$\hat{F}^*_{X|Y_j^*}(W_k^*) = \sum_{i=1}^{m^*} \hat{p}^*_{ij}\ \mathbf{I}\{W_i^* \le W_k^*\},\ j = 1, 2, \cdots, m_0^*. \tag{152}$$

Adjustment for $\hat{p}^*_{kj}$:

- For fixed Y_j^*, find $W_l^* = \max\{W_i^* \mid Z_i^* \le Y_j^*\}$.
- Then recalculate this $\hat{p}^*_{lj}$ to make $\sum_{k=1}^{l} \hat{p}^*_{kj} = 1$.

Let $b_j = G_n^*(Y_j^*)$, for $j = 1, 2, \cdots, m_0^*$. Then,

$$\hat{q}^*_{ij} = b_j\ \hat{p}^*_{ij}\ - b_{j-1}\ \hat{p}^*_{i(j-1)},\ \hat{p}^*_{j0} = 0,\quad i = 1, 2, \cdots, m^*, j = 1, 2, \cdots, m_0^*, \tag{153}$$

and

$$\hat{G}^*(x, z) = \sum_{j=1}^{m_0^*} \sum_{i=1}^{m^*} \hat{q}^*_{ij}\ \mathbf{I}\{W_i^* \le x, Y_j^* \le z\}. \tag{154}$$

Adjustment for $\hat{q}^*_{ij}$:

- Let Δ= sum of all negative $\hat{q}^*_{ij}$'s;
- Let $\Delta_t = 1 - \Delta$ = sum of all positive $\hat{q}^*_{ij}$'s;
- Rewrite

$$\hat{G}^*(x, z) \ = \ \sum_{j=1}^{m_0^*} \sum_{i=1}^{m^*} \hat{q}^{'*}_{ij}\ \mathbf{I}\{W_i^* \le x, Y_j^* \le z\},$$

where

$$\hat{q}^{'*}_{ij} \ = \ \begin{cases} 0 & \text{if } \hat{q}^*_{ij} \le 0, \\ \hat{q}^*_{ij}(1 - \frac{|\Delta|}{\Delta_t}) & \text{if } \hat{q}^*_{ij} > 0, \end{cases} \tag{155}$$

Note we still use $\hat{q}^*_{ij}$ to represent $\hat{q}^{'*}_{ij}$ for the rest of this section.

Step 9: Compute $\hat{F}_n^*(x)$ based on $\hat{G}^*$ and $\hat{\theta}^*$.

First calculate $\hat{\theta}^*$: Let

$$\phi(\theta^*) = (\bar{Z}^*) - \sum_{k=1}^{m^*} \hat{b}_k \left(\frac{\sum_{j=1}^{m_0^*} \sum_{i=k}^{m^*} \hat{q}_{ij}^* Y_j^* \exp(Y_j^* \theta^*)}{\sum_{j=1}^{m_0^*} \sum_{i=k}^{m^*} \hat{q}_{ij}^* \exp(Y_j \theta^*)} \right), \tag{156}$$

in which $\hat{b}_k = \sum_{j=1}^{m_0^*} \hat{q}_{kj}^*$, and $\bar{Z}^* = \frac{1}{n} \sum_{i=1}^{n} Z_i^*$. We use bisection algorithm to find the $\hat{\theta}^*$ as a solution of $\phi(\theta^*) = 0$, and the stopping rule is $|\phi(\hat{\theta}^*)| < 0.001$.

Then, we calculate $\hat{q}_k^*$:

$$\hat{q}_k^* \;=\; n \,(\sum_{l=1}^{m_0^*} \hat{q}_{kl}^*) \left\{ \log \frac{\sum_{i=1}^{m^*} \sum_{j=1}^{m_0^*} \mathbf{I}\{W_k^* \le W_i^*\} \exp(Y_j^* \hat{\theta}^*) \hat{q}_{ij}^* \;-\; n^{-1}}{\sum_{i=1}^{m^*} \sum_{j=1}^{m_0^*} \mathbf{I}\{W_k^* \le W_i^*\} \exp(Y_j^* \hat{\theta}^*) \hat{q}_{ij}^*} \right\}. \tag{157}$$

If $\sum_{l=1}^{m_0^*} \hat{q}_{kl}^* = 0$, then let $\hat{q}_k^* = 0$.

If $\sum_{i=1}^{m^*} \sum_{j=1}^{m_0^*} \mathbf{I}\{W_k^* \le W_i^*\} \exp(Y_j^* \hat{\theta}^*) \hat{q}_{ij}^* \;= 0$, then let $\hat{q}_k^* = 0$.

Calculate $\hat{F}_n^*$:

$$\log(\bar{\hat{F}}_n^*(t)) = \sum_{k=1}^{m^*} \hat{q}_k^* \mathbf{I}\{W_k^* \le t\}. \tag{158}$$

Therefore,

$$\hat{F}_n^*(t) = 1 \;-\; \exp\left\{ \sum_{k=1}^{m^*} \hat{q}_k^* \mathbf{I}\{W_k^* \le t\} \right\}. \tag{159}$$

which can be written as

$$\hat{F}_n^*(t) = \sum_{k=1}^{m^*} \hat{r}_k \; \mathbf{I}\{W_k^* \le t\}. \tag{160}$$

Step 10: Calculate T_n^*:

First, calculate

$$\hat{F}_n^*(V_i) = \sum_{k=1}^{m^*} \hat{r}_k \, \mathbf{I}\{W_k^* \leq V_i\}, \; i = 1, 2, \; \cdots, n, \tag{161}$$

then compute

$$T_n^* = \sqrt{n} \parallel \hat{F}_n^*(V_i) - \hat{F}_n(V_i) \parallel = \sqrt{n} \max_{1 \leq i \leq n} |\hat{F}_n^*(V_i) - \hat{F}_n(V_i)|, \tag{162}$$

whose distribution estimates that of $T_n = \sqrt{n} \parallel \hat{F}_n - F_0 \parallel$.

6.3 Simulation Results

In this section, we present some simulation results. All the figures mentioned below are listed in the Appendix. The simulation samples described in Section 6.2 have true $\beta = 1$ in (101). Here, we generate 1000 such samples.

Estimation for β: In Table 6.1, we compare the performance of $\hat{\beta}$ and $\hat{\theta}$, where the simulation s.d.'s are given in the parenthesis next to the simulation averages. Here, to compare $\hat{\beta}$, we use S-plus. The results in Table 6.1 show that $\hat{\beta}$ and $\hat{\theta}$ have very similar performance. However, $\hat{\beta}$ does not apply to complicated types of censored data, such as doubly censored data and partly interval-censored data, while our method does.

Table 6.1: Comparison of $\hat{\beta}$ and $\hat{\theta}$

	Avg. of $\hat{\beta}$	Avg. of $\hat{\theta}$	Censoring Rate
n=50	1.038 (0.232)	1.078 (0.289)	18.8%
n=100	1.010 (0.140)	1.063 (0.197)	18.9%
n=200	1.008 (0.099)	1.064 (0.145)	18.8%

Estimation of distribution function F_0: Let Exp(μ) represent the exponential distribution with mean μ, and $\hat{F}_n$ represent the estimated d.f. calculated by (144), while F_0 represent the true d.f. calculated by (145) for the sample. We generate one noncensored sample with sample size $n = 100$ from Z = Exp(1) and X = Exp(Exp(-Z)), Figure A compares $\hat{F}_n$ with F_0, where $\hat{\beta}$ is used to compute $\hat{F}_n$. Figure A.4 compares $\hat{F}_n$ with F_0, where $\hat{\theta}$ is used to compute $\hat{F}_n$. From Figures A and A.4, it is evident that two methods have little difference, and both $\hat{F}_n$'s are very good estimates for F_0.

Also, we generate one right censored sample with sample size $n = 100$ as described in Section 6.2. Figure A compares $\hat{F}_n$ with F_0, where $\hat{\beta}$ is used to compute $\hat{F}_n$, while Figure A compares $\hat{F}_n$ with F_0, where $\hat{\theta}$ is used to compute $\hat{F}_n$. These figures show that two methods differ little. Moreover, we generate one right censored sample of the same type with sample size $n = 200$. Figure A compares $\hat{F}_n$ with F_0, where $\hat{\beta}$ is used to compute $\hat{F}_n$, while Figure A compares $\hat{F}_n$ with F_0, where $\hat{\theta}$ is used to compute $\hat{F}_n$. Figures A and A show that the discrepancy of the $\hat{F}_n$ from the true d.f. F_0 is getting smaller as the sample size gets larger. Again, there is very little difference between using $\hat{\beta}$ or $\hat{\theta}$ to compute $\hat{F}_n$. However, our method is easy to compute and applicable to complicated type of censored data, such as doubly censored data and partly interval-censored data.

Simulation distributions of statistics T_n *and* T_n^**:* Here we have $T_n = \sqrt{n} \parallel \hat{F}_n - F_0 \parallel$ and $T_n^* = \sqrt{n} \parallel \hat{F}_n^* - \hat{F}_n \parallel$. For 1000 generated samples with sample size $n = 100$, we generate one bootstrap sample for each sample. Then, statistics $T_n = \sqrt{n} \parallel \hat{F}_n - F_0 \parallel$ and $T_n^* = \sqrt{n} \parallel \hat{F}_n^* - \hat{F}_n \parallel$ are calculated for each sample and each bootstrap sample as in (146) and (162), respectively. $\hat{F}_n$ is calculated using $\hat{\theta}$ by our method as in (144), F_0 is calculated as in (145), and $\hat{F}_n^*$ using $\hat{\theta}^*$ by our method as in (160). Figure A in the Appendix displays the simulate distributions of T_n and T_n^*, which shows that T_n^* provides good estimate for T_n. It should be noted that it is not practical to use $\hat{\beta}$ for computing $\hat{F}_n$ when bootstrap method is used for T_n^*. Thus, in our simulation studies we only considered the use of $\hat{\theta}$ here.

CHAPTER 7

CONCLUDING REMARKS

From our simulation results, it is shown that semi-parametric empirical likelihood method is powerful in hypothesis tests for two sample problems on Cox model and stratified Cox model, and especially useful for complicated types of censored data, like right censored, doubly censored and partly interval-censored data.

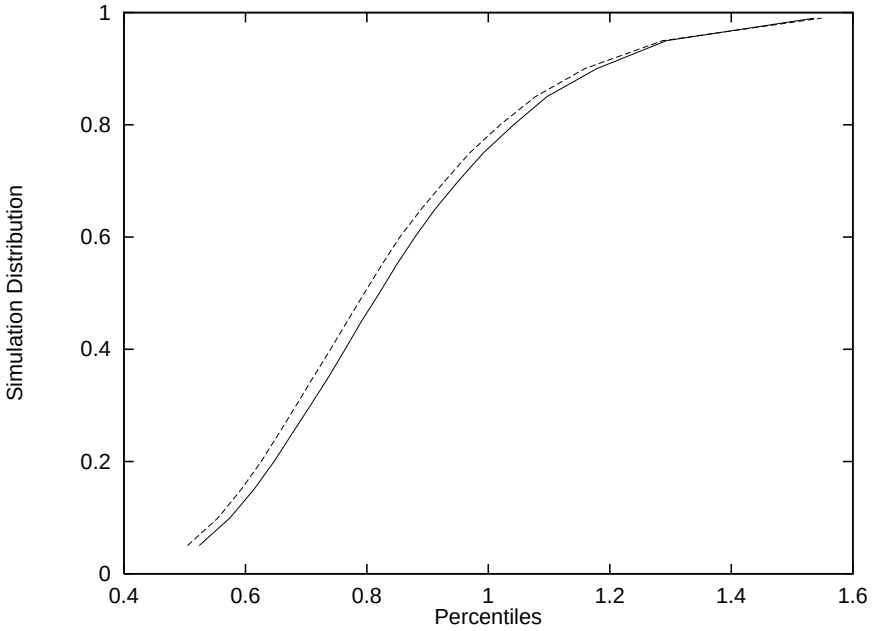
Our proposed approach is computationally simple. Along with the construction of the test, we provide a consistent semi-parametric maximum likelihood estimator for β_0 under model assumption (61) for two sample Cox model. It should be noted that all results here actually hold for any censored data whose NPMLE for the distribution function is asymptotically Gaussian, and our method presented here can be easily extended to $k-$sample Cox model.

For stratified Cox model, we proposed a new approach to estimate the parameter β under model assumption (102) which applies for complicated types of censored data. We also constructed the goodness-of-fit test. Our simulation results show that our method is as good as the usual Newton-Rahpson method, but our method also applies for the complicated types of censored data while the usual Newton-Rahpson method can not.

For stratified Cox model, our simulation results for different sample sizes are not stable. Further studies on the computation issues related to this problem are to be conducted.

APPENDIX A

FIGURES



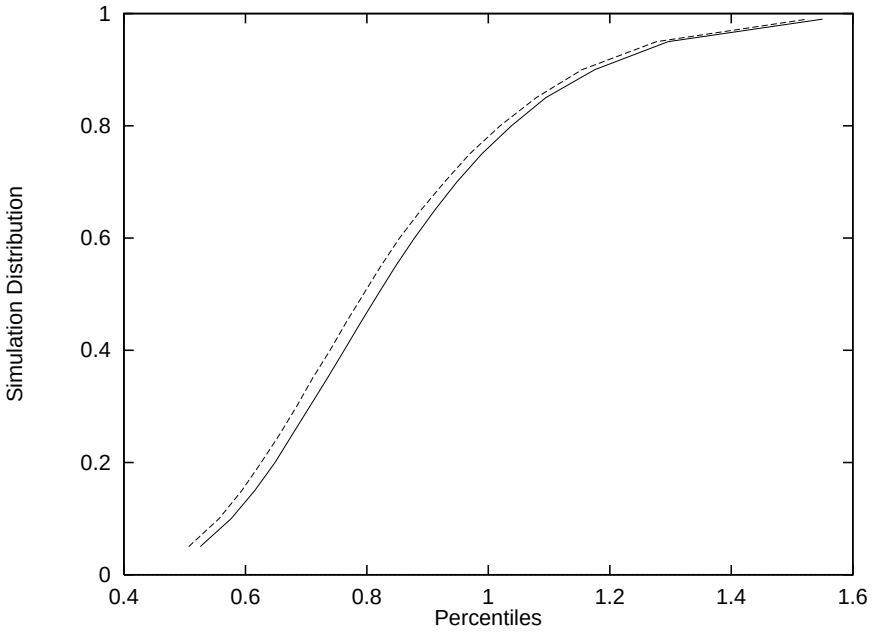
T_n = solid line; T_n^* = dashed line.

Simulation loops: 20000.

Noncensored sample for $n_0 = 100 : G = \exp(1)$; $n_1 = 150 : H = \exp(0.5)$;

[average of $\tilde{\eta}$] = 2.026 with s.d. = 0.285.

Figure A.1: **Two-Sample simulation for Noncensored Samples 1**



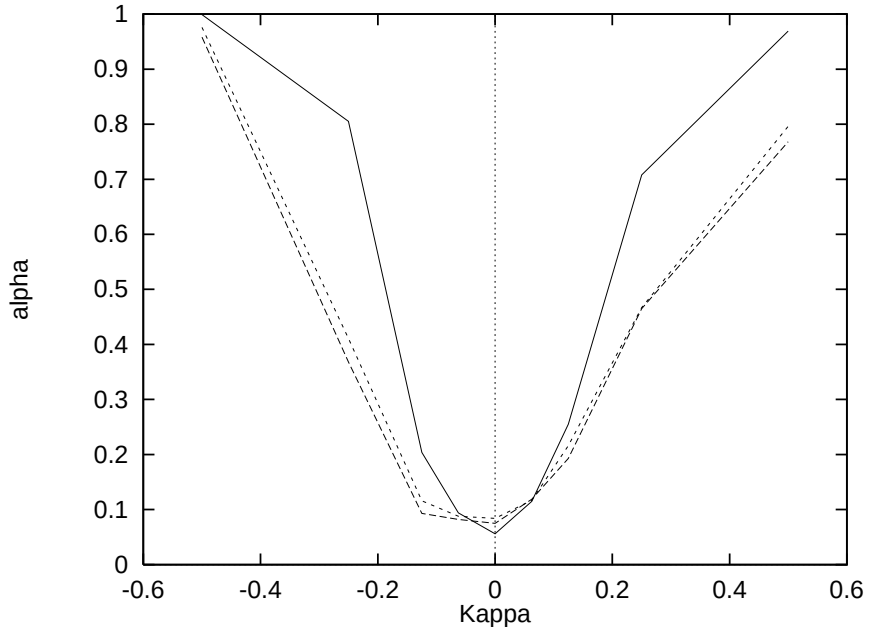
T_n = solid line; T_n^* = dashed line.

Simulation loops: 20000.

Noncensored sample for $n_0 = 150 : G = \exp(1)$; $n_1 = 100 : H = \exp(0.5)$;

[average of $\tilde{\eta}$] = 2.031 with s.d. = 0.282.

Figure A.2: **Two-Sample simulation for Noncensored Samples 2**



Noncensored: Solid line, right censored: long dashed line, doubly censored: short dashed line.

$G_0 = \exp(1)$ vs. $H_0 = \exp(0.5) + \kappa U$

Noncensored sample for $n_0 = 150$ and $n_1 = 100$;

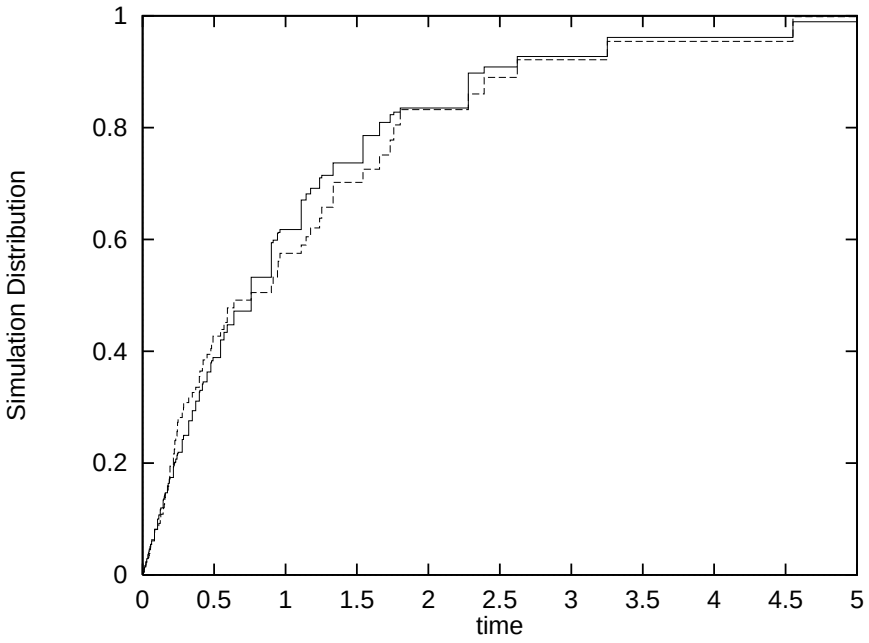
Right censored sample for $n_0 = 150$ and $n_1 = 100$: $C_G = \exp(2)$ (33.1% censoring); $C_H = \exp(1)$ (33.1%censoring);

Doubly censored sample for $n_0 = 150$ and $n_1 = 100$:

$C_G = \exp(3)$ (24.9% censoring); $D_G = (2/3)C_G - 2.50$ (19.0% censoring);

$C_H = \exp(1)$ (33.1% censoring); $D_H = (2/3)C_H - 2.50$ (1.5% censoring);

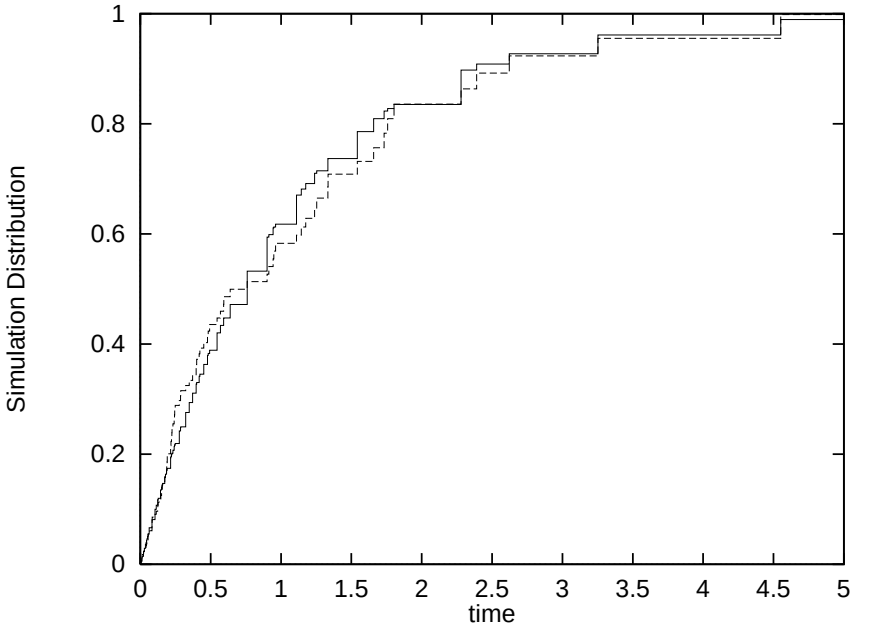
Figure A.3: **Power of Tests with 95% Significance Level**



F_0 = solid line; $\hat{F}_n$ = dashed line (using $\hat{\beta}$).

Noncensored sample for $n = 100$, $\hat{\beta} = 1.062$ (from S-plus);

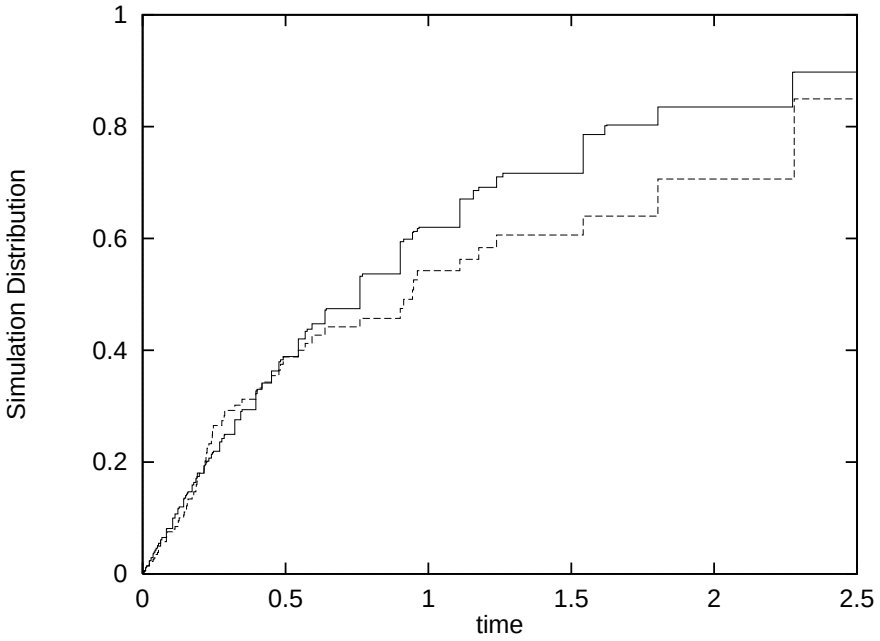
Figure A.4: **Stratified Cox Model with Noncensored Samples 2**



F_0 = solid line; $\hat{F}_n$ = dashed line (using $\hat{\theta}$).

Noncensored sample for $n = 100$, $\hat{\theta} = 1.031$;

Figure A.5: **Stratified Cox Model with Noncensored Samples 1**

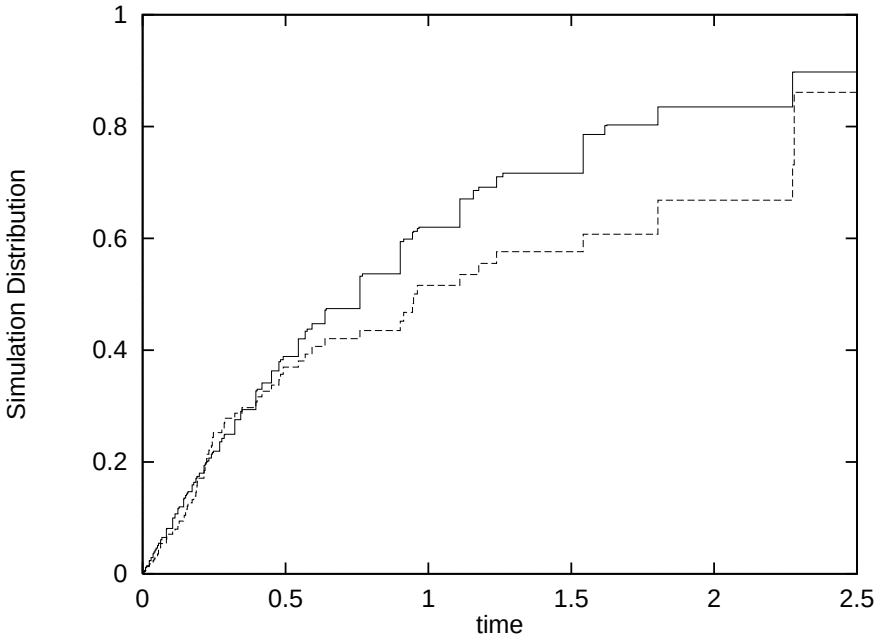


F_0 = solid line; $\hat{F}_n$ = dashed line (using $\hat{\beta}$).

Right censored sample for $n = 100$;

$\hat{\beta} = 1.065$ (from S-plus), censored observations 16.0%;

Figure A.6: **Stratified Cox Model with Right Censored Samples 1**

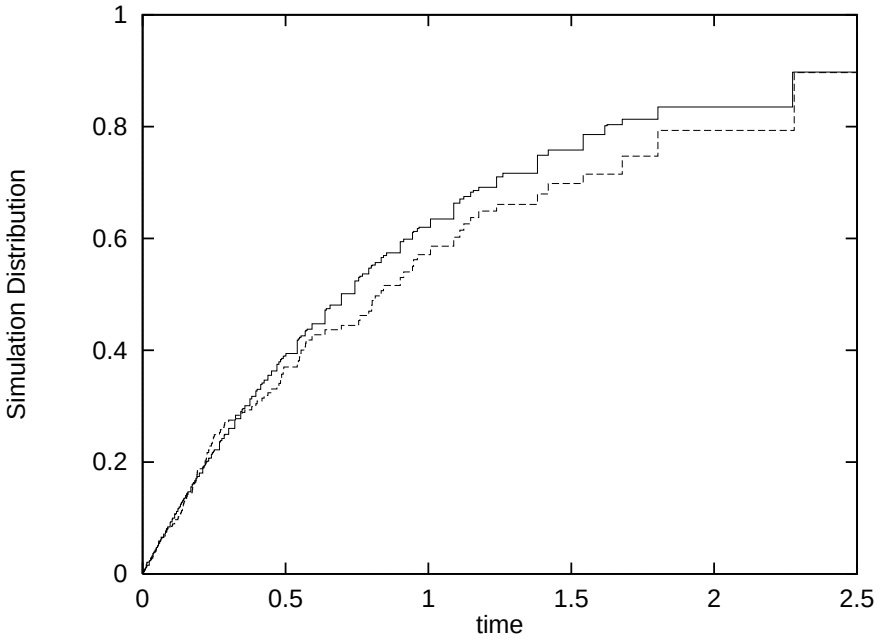


F_0 = solid line; $\hat{F}_n$ = dashed line (using $\hat{\theta}$).

Right censored sample for $n = 100$;

$\hat{\theta} = 1.096$, censored observations 16.0%;

Figure A.7: **Stratified Cox Model with Right Censored Samples 2**

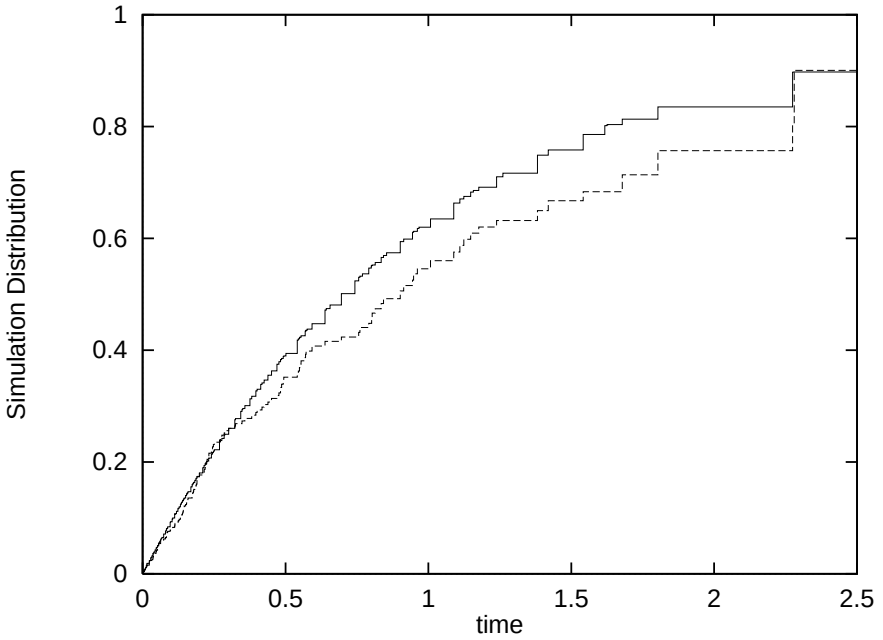


F_0 = solid line; $\hat{F}_n$ = dashed line (using $\hat{\beta}$).

Right censored sample for $n = 200$;

$\hat{\beta} = 1.022$ (from S-plus), censored observations 17.0%;

Figure A.8: **Stratified Cox Model with Right Censored Samples 3**

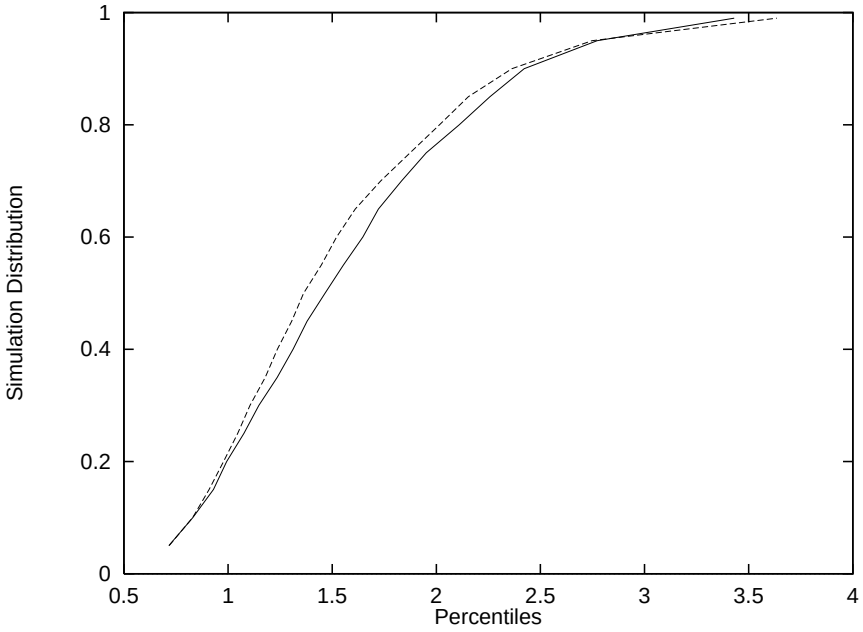


F_0 = solid line; $\hat{F}_n$ = dashed line (using $\hat{\theta}$).

Right censored sample for $n = 200$;

$\hat{\theta} = 1.046$, censored observations 17.0%;

Figure A.9: **Stratified Cox Model with Right Censored Samples 4**



T_n = solid line; T_n^* = dashed line.

Simulation loops: 1000.

Right censored sample for $n = 100$, censored observations 18.9%.

Figure A.10: **Stratified Cox model T_n vs. T_n^***

REFERENCES

[1] Bickel, P.J. and Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.*, Vol. 9, 1196-1217.

[2] Bickel, P.J. and Ren, J. (1996). The m out of n bootstrap and goodness of fit tests with doubly censored data. *Robust Statistics, Data Analysis and Computer Intensive Methods*. Lecture Notes in Statistics, Springer Verlag, Vol. 109, 35-47.

[3] Casella, G. and Berger R.L. (1990). *Statistical Inference*. Wadsworth & Brooks/Cole Advanced Books & Software.

[4] Chang, M.N. (1990). Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *Ann. Statist.*, Vol. 18, 391-404.

[5] Chang, M.N. and Yang, G.L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *Ann. Statist.*, Vol. 15, 1536-1547.

[6] Cox, D.R. (1972). Regression models and life-tables (with discussion). *Journal of Royal Statistical Society, Series B,* Vol. 34, 187-220.

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

[8] DiCiccio, T.J., Hall, P., and Romano, J. (1991). Empirical likelihood is Bartlett-correctable. *Ann. Statist.*, Vol. 19, 1053-1061.

[9] Efron, B. (1967). The two sample problem with censored data. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 4, 831-853.

[10] Efron, B. (1979). Bootstrap methods: Another look at jackknife. *Ann. Statist.*, Vol. 7, No. 1, 1-26.

[11] Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.

[12] Enevoldsen, A.K., Borch-Johnson, K., Kreiner, S., Nerup, J., and Deckert, T. (1987). Declining incidence of persistent proteinuria in type I (insulin-dependent) diabetic patient in Denmark. *Diabetes*, Vol. 36, 205-209.

[13] Gastrointestinal Tumor Study Group: Schein, P.D., Bruckner, H.W., Douglass, H.O., Mayer, R. *et al.* (1982). A comparison of combination chemotherapy and combined modality therapy for locally advanced gastric carcinoma. *Cancer*, Vol. 49, 1771-1777.

[14] Gill, R. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.*, Vol. 11, 49-58.

[15] Gill, R. and Schumacher, M. (1987). A simple test of the proportional hazards assumption. *Biometrika*, Vol. 74, 289-300.

[16] Giné, E. and Zinn, J. (1990). Bootstraping general empirical measures. *Ann. Prob.*, Vol. 18, 852-869.

[17] Gu, M.G. and Zhang, C.H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann. Statist.*, Vol. 21, No. 2, 611-624.

[18] Hosmer, D.W. and Lemeshow, S. (1999). *Applied Survival Analysis–Regression Modeling of Time to Event Data*. Wiley Series in Probability and Statistics.

[19] Huang, J. (1999). Asymptotic properties of nonparametric estimation based on partly interval-censored data. *Statistica Sinica*, Vol. 9, 501-519.

[20] Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, Vol. 53, 457-481.

[21] Kleinbaum, D.G. (1995). *Survival Analysis*. Springer.

[22] Knight, K. (2000). *Mathematical Statistics*. Chapman & Hall/CRC.

[23] Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data*, Second Edition. Wiley.

[24] Leiderman, P.H., Babu, D., Kagia, J., Kraemer, H.C., and Leiderman, G.F. (1973). African infant precocity and some social influences during the first year. *Nature*, Vol. 242, 247-249.

[25] Lin, D.Y. (1991). Goodness-of-fit analysis for the Cox regression model based on a class of parameter estimators. *J. Amer. Statist. Assoc.*, Vol. 86, 725-728.

[26] Mykland, P.A. (1995). Dual likelihood. *Ann. Statist.*, Vol. 23, 396-421.

[27] Mykland, P.A. and Ren, J. (1996). Algorithms for computing self-consistent and maximum likelihood estimators with doubly censored data. *Ann. Statist.*, Vol. 24, 1740-1764.

[28] 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*, Vol. 48, 951-959.

[29] Owen, A.B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, Vol. 75, No. 2, 237-249.

[30] Owen, A.B. (1990). Empirical likelihood ratio confidence regions . *Ann. Statist.*, Vol. 18, 90-120.

[31] Owen, A.B. (1991). Empirical likelihood for linear models. *Ann. Statist.*, Vol. 19, 1725-1747.

[32] Owen, A.B. (2001). *Empirical Likelihood*. Chapman & Hall, New York.

[33] Peer, P.G.M., Van Dijck, J.A., Hendriks, J.H., Holland, R., and Verbeek, A.L. (1993). Age-dependent growth rate of primary breast cancer. *Cancer*, Vol. 71, 3547-3551.

[34] Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.*, Vol. 22, No. 1, 300-325.

[35] Ren, J. (1995). Generalized Cramér-von Mises tests of goodness of fit for doubly censored data. *Ann. Inst. Statist. Math.*, Vol. 47, 525-549.

[36] Ren, J. and Gu, M.G. (1997). Regression M-estimators for doubly censored data. *Ann. Statist.*, Vol. 25, 2638-2664.

[37] Ren, J. and He, B. (2005). Goodness-of-fit tests for the Cox model with doubly censored data or partly interval-censored data. (Submitted to *Canadian Journal of Statistics*).

[38] Ren, J. and Peer, P.G.M. (2000). A study on effectiveness of screening mammograms. *International J. of Epidemiology*, Vol. 29, 803-806.

[39] Ren, J., Su, X., and He, B. (2006). Comparing baseline hazard functions in stratified Cox models with censored data. (in preparation).

[40] Smith, P.J. (2002). *Analysis of Failure and Survival Data*. Chapman & Hall/CRC.

[41] Stute, W. and Wang, J.L. (1993). The strong law under random censorship. *Ann. Statist.*, Vol. 21, 1591-1607.

[42] Thomas, D.R. and Grunkemeier, G.L. (1975). Confidence interval estimation of survival probabilities for censored data. *J. Amer. Statist. Assoc.*, Vol. 70, 865-871.

[43] Turnbull, B.W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Amer. Statist. Assoc.*, Vol. 69, 169-173.

[44] Wilks, S.S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.*, Vol. 9, 60-62.

[45] Yang, S. and Prentice, R. (2005). Semiparametric analysis of short-term and long-term hazard ratios with two-sample survival data. *Biometrika*, Vol. 92, 1-17.