Inferences in the Interval Censored Exponential Regression Model

Inferences in the Interval Censored Exponential Regression Model

By

Defen Peng

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Master of Science

McMaster University

MASTER OF SCIENCE (2009) McMaster University

(Statistics) Hamilton, Ontario

TITLE: Inferences in the Interval Censored Exponential Regression Model

AUTHOR: Defen Peng, B.Sc. (McMaster University)

SUPERVISOR: Dr. Román Viveros-Aguilera

NUMBER OF PAGES: x, 87

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Román Viveros-Aguilera. For his continual help, experience sharing and professional advice. I will be eternally grateful.

My special thanks are indebted to Dr. Angelo Canty and Dr. Rong Zhu, who looked closely at the final version of the thesis for comments, questions, as well as English style and grammar, and offered suggestions for improvement. I was very impressed in their preciseness and graveness.

I also would like to acknowledge all of the faculty members in Statistics for their teaching and instruction during the years I spent at McMaster University.

Many thanks also go to Professor Gilbert MacKenzie, Centre of Biostatistics, University of Limerick, Ireland (www.ul.ie/biostatistics) for his generous support of references and helpful suggestions.

Finally, I would like to thank my family and friends for their constant encouragement and support.

Contents

List of Tables

List of Figures

Abstract

The problem of estimation when the data are interval censored has been investigated by several authors. Lindsey and Ryan (1998) considered the application of conventional methods to interval mid (or end) points and showed that they tended to underestimate the standard errors of the estimated parameters and give potentially misleading results. MacKenzie (1999) and Blagojevic (2002) conjectured that the estimator of the parameter was artificially precise when analyzing inspection times as if they were exact when the "time to event" data followed an exponential distribution. In this thesis, we derive formulae for pseudo and true (or exact) likelihoods in the exponential regression model in order to examine the consequences for inference on parameters when the pseudo-likelihood is used in place of the true likelihood. We pay particular attention to the approximate bias of the maximum likelihood estimates in the case of the true likelihood. In particular we present analytical work which proves that the conjectures of Lindsey and Ryan (1998), MacKenzie (1999) and Blagojevic (2002) hold, at least for the exponential distribution with categorical or continuous covariates.

We undertake a simulation study in order to quantify and analyze the relative performances of maximum likelihood estimation from both likelihoods. The numerical

evidence suggests that the estimates from true likelihood are more accurate. We apply the proposed method to a set of real interval-censored data collected in a Medical Research Council (MRC, UK) multi-centre randomized controlled trial of teletherapy in the age-related macular disease (the ARMD) study.

Chapter 1

Introduction

Interval censoring occurs mainly in longitudinal studies and clinical trials, where, repeated measurements are taken on individuals, usually at pre-specified inspection times. In these studies it is often the case that some individuals may develop the event of interest between two inspection times, say U and V. In such a case the exact time to event, t, is unknown, i.e., $U < t \leq V$, which is called interval censoring.

Three special cases of interval-censored data have been well-studied: (a) if $U = 0$, we have left-censored survival data; (b) if $V = \infty$, we have right-censored survival data, and three types of right censoring, Type I censoring, Type II censoring and independent random censoring were shown in Lawless (2003); (c) if either $U = 0$ or $V = \infty$, the data are usually referred to as current status data (Kalbfleisch and Prentice (1980)). Left censoring is rare and is not considered in this thesis, also we do not consider the current status data. The interval censored (IC) and right censored (RC) data are of interest.

Interest in interval censored data analysis has increased recently because the data involved in HIV/AIDS studies are often interval censored (Satten (1996) and Hart et al. (2002)). In practice, clinicians (e.g., Bergink et al. (1998)) utilize the mid or end-point, as if they were actual times to event. The common method of analyzing such data is to assume that the event occurred at the end (or midpoint) of each interval and then to apply standard survival methods for time to event data (Turnbull (1976), Lindsey and Ryan (1998)). Application of conventional methods to interval mid (or end) points may tend to underestimate the standard errors of the estimated parameters and give potentially misleading results (Lindsey and Ryan (1998)). Also, an investigation of the exponential distribution showed that the estimator of the parameter was artificially precise when analyzing inspection times as if they were exact (MacKenzie (1999) and Blagojevic (2002)).

This thesis is concerned with various aspects of the analysis of interval censored data, but the main focus is on the precision of the resulting estimators.

1.1 Aims

In this thesis, we consider the exponential regression model and investigate the consequences of inference on parameters when a pseudo-likelihood is used instead of the true likelihood to interval censored data. Such data arise often in longitudinal studies such as randomized controlled trials where the regression parameter measures the effect of treatment and is then of special interest.

The main goal of this thesis is to compare the precision of the estimators, especially, the regression parameters, in an exponential regression model when using the correct interval-censored likelihood and a pseudo-likelihood.

More specifically, we aim to

1. identfy, and interpret any clear structural differences or similarities that may arise between the two likelihoods and in different quantities derived from them.

2. study the use of first term approximation of the conditional odds on the survival in the case of the true likelihood.

3. conduct simulation to obtain numerical comparison between estimates from two likelihoods methods. The comparison is regarding to means and mean square errors of the estimators.

4. apply to a set of real interval-censored data arising in a MRC randomized clinical trial of teletherapy in age-related macular degeneration (ARMD).

1.2 Overview

In Chapter 2, we brief the theory for analysing interval-censored data using survival distributions to model the time to failure. Chapter 3 derives the forms of maximum likelihood estimates from the true likelihood and pseudo-likelihood respectively, which will be used to fit the exponential regression distribution to interval-censored data. In Chapter 4, we show the estimate from pseudo-likelihood is artificially precise when compared to the true likelihood; and we find the form of the Fisher information for the interval-censored data. In Chapter 5, we first make a simulation-based estimation from true likelihood and pseudo-likelihood for interval censored data, and analyze the results. Finally, we apply the proposed method to a real data obtained from an age-related macular disease macular disease (ARMD) study (Hart et al. (2002)) in Chapter 6. Concluding remarks and future study are summarized in Chapter 7.

Chapter 2

Review of Basic Survival Analysis

2.1 General Concepts

Let T be a continuous random variable representing the time to failure, so $T \geq 0$. The probability density function (pdf) $f(t)$ of T is defined via

$$\mathrm{P}\{t < T < t + \delta t\} = \int_t^{t+\delta t} f(u)\mathrm{d}u \simeq f(t)\delta t$$

for $t \geq 0$, and the cumulative distribution (cdf), F of T, which is based on aggregating probability, is given by

$$F(t) = \mathrm{P}\{T \leq t\} = \int_0^t f(u)\mathrm{d}u,$$

again for $t \geq 0$.

Any distribution in survival analysis can also be characterised by its survival function, $S(t)$, and the hazard function, $\lambda(t)$. The survival function is the probability that an individual survives at least time t, so that

$$S(t) = 1 - F(t)$$

for $t \geq 0$. The hazard (or instant hazard) function is defined in terms of the probability that, given its survival until t, an item then fails in the interval $(t, t+\delta t]$. Since this probability is, via the conditioning argument,

$$\begin{aligned} P(t < T \leq t + \delta t | \; T > t) &= \frac{P(t < T \leq t + \delta t)}{P(T > t)} \\ &= \frac{S(t) - S(t + \delta t)}{S(t)} \\ &\simeq \frac{f(t)\delta t}{S(t)}, \end{aligned}$$

we see that

$$\begin{aligned} \lambda(t) &= \lim_{\delta t \to 0} \frac{\{S(t) - S(t + \delta t)\}/S(t)}{\delta t} \\ &= \lim_{\delta t \to 0} \left\{ \frac{F(t + \delta t) - F(t)}{\delta t} \right\} \frac{1}{S(t)} \\ &= \frac{f(t)}{S(t)}. \end{aligned}$$

Thus the hazard function is sometimes called a "conditional failure rate", since the denominator reflects the conditional probability given survival to time t. The cumulative hazard function is given by

$$\Lambda(t) = \int_0^t \lambda(u) \mathrm{d}u,$$

from which

$$S(t) = \exp\{ - \Lambda(t)\} = \exp\left\{ - \int_0^t \lambda(u) \mathrm{d}u \right\}. \tag{2.1.1}$$

In survival analysis, estimation and inference of the survival function $S(t)$ at each time t are of interests. The method for implementing this depends on whether the form of the probability density function (pdf) $f(t)$ or cumulative function $F(t)$ is known or not.

2.2 Non-parametric Method

In practice, we must have estimators that can be used for censored data and, in particular, with right censored data and unknown form of the distribution. For this purpose, the *Kaplan-Meier estimator* (refer to Lawless (2003)) of the survival function is a popular and useful method.

Suppose that from n independent units on test, the r distinct failure times $t_{(1)} < t_{(2)} < \cdots < t_{(r)}$ $(r \leq n)$ are observed. At time $t_{(j)}$, n_j units are "at risk", that is, in operation, and d_j units fail before time $t_{(j)}$ immediately. This number n_j covers all the units that are observed for a duration at least $t_{(j)}$, including those that will subsequently fail as well as those that will subsequently be censored.

We will now estimate the survival function $S(t)$ at each time $t_{(i)}$ as follows,

$$\begin{aligned} S(t_{(i)}) &= P(T > t_{(i)}) \\ &= P(T > t_{(1)})P(T > t_{(2)}|\, T > t_{(1)}) \cdots P(T > t_{(i)}|\, T > t_{(i-1)}). \end{aligned}$$

This is derived by re-expressing, for example, the probability of the event $\{T > t_{(2)}\}$ using the basic identity in conditional probability, $P(A \cap B) = P(A)P(B|A)$, with A as the event $\{T > t_{(1)}\}$ and B as the event $\{T > t_{(2)}\}$.

The simple estimator of $P(T > t_{(1)})$ is,

$$P(T > t_{(1)}) = 1 - p_1 = 1 - \frac{d_1}{n_1} = \frac{n_1 - d_1}{n_1},$$

where p_1 is the relative frequency of failures in the interval $[0, t_{(1)})$ and n_1 is the number of units that are at risk just before time $t_{(1)}$. Then, by the same logic, the conditional probability for failure in the next interval $[t_{(1)}, t_{(2)})$ is

$$P(T > t_{(2)}|\, T > t_{(1)}) = 1 - p_2 = 1 - \frac{d_2}{n_2} = \frac{n_2 - d_2}{n_2},$$

where p_2 is the relative frequency of failures in the interval $[t_{(1)}, t_{(2)})$ and n_2 is the number of units that are at risk just before time $t_{(2)}$, and so on. Proceeding in this way, we finally obtain

$$\begin{aligned}\hat{S}(t) &= \frac{n_1 - d_1}{n_1} \times \frac{n_2 - d_2}{n_2} \times \cdots \times \frac{n_i - d_i}{n_i} \quad \text{for } t : t_{(i)} \le t < t_{(i+1)} \\ &= \prod_{j:\ t_{(j)} \le t} \frac{n_j - d_j}{n_j}.\end{aligned}$$

This is the *Kaplan-Meier estimator* of the survival function. Observe that this is a *non-parametric estimator* since it does not assume any particular functional form for the curve $S(t)$. This makes the estimator important and useful.

Since $\hat{S}(t)$ is a sample statistic, it is desirable to have an estimate of its standard deviation, called the *standard error.* The formula usually employed for this purpose is the so-called Greenwood's formula [1] (refer to Lawless (2003)) given by

$$\widehat{SE}\big(\hat{S}(t)\big) = \hat{S}(t) \left\{ \sum_{t_{(j)} \le t} \frac{d_j}{n_j(n_j - d_j)} \right\}^{1/2}.$$

2.3 Parametric Models

The estimators introduced in Section 2.2 are *non-parametric estimators*, since we did not assume any particular functional form for the survival or hazard functions. This freedom from assumptions is often very desirable. On the other hand, it may be better to use the functional form when we know what it is. Typically, parametric models tend to have smaller standard errors when estimating quantities such as the median and hazards than models without a specified distribution.

[1] *Major Greenwood (1880-1949) was Professor of Epidemiology and Vital Statistics at the London School of Hygiene and Tropical Medicine and also worked on distribution theory for accident data.*

In estimating the parameters of statistical distributions so that they can be fitted to the survival data, several methods have been established, such as the method of moments, the maximum likelihood estimation, the least squares method, and the Bayesian method. See Casella and Berger (2001) for an overview of estimation methods. However, the emphasis has fallen heavily on maximum likelihood estimation.

The method of maximum likelihood is, by far, the most popular technique for deriving estimators. The likelihood function can be used to examine the whole range of possible parameter values, and determine which values are most consistent (or, in common parlance, "likely") with respect to the data (complete or censored). There are a lot of different statistical distributions that have been found to be most useful for describing survival data. These have also been used in the literature as base models to build more general models possessing a lot of flexibility to fit different forms of data. The following are a number of parametric distributions commonly used.

2.3.1 Exponential Distribution

The probability density function of the exponential distribution[2] is

$$f(t) = \lambda \mathrm{e}^{-\lambda t}, \quad t > 0,\ \lambda > 0.$$

As seen in Section 2.1, the survival function is

$$S(t) = \mathrm{e}^{-\lambda t}, \quad t > 0,\ \lambda > 0,$$

and the hazard function is

$$\lambda(t) = \frac{f(t)}{S(t)} = \lambda.$$

[2] *The exponential distribution occurs naturally when describing the lengths of the inter-arrival times in a homogeneous Poisson process.*

The expected value and variance of the survival time T are then readily obtained to be

$$E(T) = \frac{1}{\lambda} \quad \text{and} \quad V(T) = \frac{1}{\lambda^2},$$

respectively.

2.3.2 Weibull Distribution

The Weibull[3]distribution is one of the most used lifetime distribution models. It is a versatile distribution that can take on characteristics of other types of distributions based on the value of its parameter γ. The probability density function of the Weibull distribution is given by

$$f(t) = \lambda\gamma t^{\gamma-1}\exp(-\lambda t^{\gamma}), \quad t > 0.$$

The probability density function has two parameters both of which are greater than zero. If $\gamma = 1$, the density function becomes $\lambda e^{-\lambda t}$ and then the survival times have an exponential distribution. The shape of the density function depends on γ which is known as the shape parameter, while λ is known as the scale parameter. The survival

[3] *"It is named after Waloddi Weibull who described it in detail in 1951, although it was first identified by Fréchet (1927) and first applied by Rosin and Rammler (1933) to describe the size distribution of particles"(Wikipedia (2009)).*

function is

$$
\begin{aligned}
S(t) &= \int_t^{\infty} \lambda\gamma u^{\gamma-1}\exp(-\lambda u^{\gamma})du \\
&= \int_{(\lambda u^{\gamma})}^{\infty} e^{-v}dv \qquad \left[\text{setting } v = \lambda u^{\gamma}\right] \\
&= \left[-e^{-v}\right]_{(\lambda t^{\gamma})}^{\infty} \\
&= \exp\left\{-\lambda t^{\gamma}\right\}
\end{aligned}
$$

and the hazard function is, therefore,

$$\lambda(t) = \lambda\gamma t^{\gamma-1}, \quad t > 0.$$

For $\gamma \neq 1$, the hazard function increases and decreases monotonically.

The first and second moments of the survival time T are then readily obtained to be

$$E(T) = \lambda^{-\frac{1}{\gamma}}\Gamma\left(1+\frac{1}{\gamma}\right) \quad \text{and} \quad E(T^2) = \lambda^{-\frac{2}{\gamma}}\Gamma\left(1+\frac{2}{\gamma}\right)$$

which yields the variance of the survival time T to be

$$V(T) = \lambda^{-\frac{2}{\gamma}}\left[\Gamma\left(1+\frac{2}{\gamma}\right) - \left\{\Gamma\left(1+\frac{1}{\gamma}\right)\right\}^2\right],$$

where $\Gamma(\cdot)$ is the gamma function.

2.3.3 Gumbel Distribution

The Gumbel distribution[4] has the pdf

$$f(t) = \frac{1}{\sigma}\exp\left\{(t-\mu)/\sigma\right\}\exp\left[-\exp\left\{(t-\mu)/\sigma\right\}\right], \quad -\infty < t < \infty,$$

[4] *One of the first scientists to apply the theory was a German mathematician Emil Gumbel (1891-1966). Gumbel's focus was primarily on applications of extreme value theory to engineering problems.*

where μ is the location parameter and σ is the scale parameter. The survival function is

$$S(t) = \exp\Big[-\exp\{(t-\mu)/\sigma\}\Big]$$

and so the hazard function is

$$\lambda(t) = \frac{1}{\sigma}\exp\Big\{(t-\mu)/\sigma\Big\}$$

which is clearly monotonously increasing.

The expected value and variance of T can be shown to be

$$E(T) = \mu - \gamma\sigma \qquad \text{and} \qquad V(T) = \frac{\pi^2}{6}\sigma^2,$$

respectively, where $\gamma = 0.5772...$ is the Euler's constant given by

$$\gamma = -\Gamma'(1) = -\int_0^\infty \mathrm{e}^{-x}\ln x\mathrm{d}x.$$

The special case with $\mu = 0$ and $\sigma = 1$ is called the standard extreme value distribution.

The greater importance of the Gumbel distribution lies in its relationship to the Weibull distribution. It can be shown that

$$T \sim Weibull(\lambda, \gamma) \qquad \Leftrightarrow \qquad \ln T \sim Gumbel(\mu, \sigma),$$

where $\mu = \log(\lambda)$ and $\sigma = 1/\gamma$.

2.3.4 Gompertz Distribution

Applications of the Gompertz[5] distribution are most notable in the analysis of mortality and actuarial data. Its simplest description is through the hazard function

[5] *Benjamin Gompertz (1779-1865), was a self educated mathematician who was elected a Fellow of the Royal Society in 1819.*

given by

$$\lambda(t) = \exp(\lambda + \gamma t)$$

which is decreasing for $\gamma < 0$, increasing for $\gamma > 0$, and constant for $\gamma = 0$ (in which case it simply reduces to the exponential distribution).

In this thesis the exponential distribution is considered. Models in which covariates have a multiplicative effect on the hazard function play a prominent role in the analysis of survival data (Lawless (2003)), hence, we will consider proportional hazards (PH) exponential regression model. Comparison with semiparametric PH model is also considered.

In practice, we can obtain the Kaplan-Meier estimate $\hat{S}(t)$ and plot it against t or a suitable function of t, in order to see if it takes nearly the same shape as assumed function $S(t)$ (Collett (2003)). We will show this in Section 6.1.

2.4 Some Formulas

We will use some well-known formulas in calculus and algebra in this thesis.

1. Taylor expansions, refer to Abramowitz and Stegun (1970).

The Taylor expansion of a real or complex function $f(x)$ that is infinitely differentiable in a neighbourhood of a real or complex number a, is the power series

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \cdots$$

which in a more compact form can be written as

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n,$$

where $n!$ denotes the factorial of n and $f^{(n)}(a)$ denotes the n-th derivative of f evaluated at the point a; the zero-th derivative of f is defined to be f itself and $(x-a)^0$ and $0!$ are both defined to be 1.

In the particular case where $a = 0$, the series is also called a Maclaurin series. Especially, the Taylor series for the exponential function e^x at $a = 0$ is

$$e^x = 1 + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \cdots$$

2. Block partitions, refer to Barndorff-Nielsen (1988).

Consider an $m \times m$ matrix M, partitioned into blocks as

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

where A is $p \times p$ and D is $q \times q$. Suppose $|A| \neq 0$.

We then have

(a)

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |A||D - BA^{-1}C| .$$

(b) If M is symmetric (whence $C = B'$) then

$$M^{-1} = \begin{bmatrix} A^{-1} + FE^{-1}F' & -FE^{-1} \\ -E^{-1}F' & E^{-1} \end{bmatrix},$$

where

$$E = D - B'A^{-1}B, \ F = A^{-1}B.$$

(c) If $|D| \neq 0$ then

$$(A + BDB')^{-1} = A^{-1} - A^{-1}B(B'A^{-1}B + D^{-1})^{-1}B'A^{-1}.$$

Chapter 3

Interval Censoring (IC)

3.1 The Interval Censored Likelihood

A longitudinal study is a correlational research study that involves repeated observations of the same items over long periods of time, then, longitudinal studies are observational. Especially, the cohort studies, a form of longitudinal study, sample a cohort, defined as a group experiencing some event (typically birth) in a selected time period, and studying them at intervals through time. Thus, in such a setting we often look at time to some event of interest, correspondingly, these times are interval censored. Consider a schematic set-up for longitudinal studies as follows

$$\begin{array}{ccccccccc} t_0^* & t_1^* & t_2^* & t_3^* & t_{k-1}^* & t_k^* & t_{m-1}^* & t_m^* & t_{m+1}^* = \infty \\ |----|----|----|\cdots\cdots|----|\cdots\cdots|----|----|----| , \end{array} \tag{3.1.1}$$

where $t_1^*, t_2^*, \cdots, t_m^*$ are inspection times and t_0^* is the baseline time. Suppose that n individuals with failure times in intervals following a continuous distribution with density function $f(t;\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a parameter vector. Denote by T_i=failure time of individual i and treat it as a random variable, and t_i=exact failure time (not observed), $i = 1, 2, \cdots, n$.

In this scheme we assume that any patient is either IC or RC, just two possibilities when we analyzing the data at the end of this study. If a patient is IC, we observe the particular interval $(t^*_{j-1}, t^*_j]$, $j = 1, 2, \cdots, m$, where her/his failure time occurred; if the patient is RC, we only know that her/his failure time occurred beyond t^*_m. Usually, the RC failure time will be a scheduled time point, therefore, this scheme should be extended when the RC occurs before t^*_m. Also we assume that the censoring is non-informative.

For the i-th patient, define the vector

$$\mathbf{w}_i = (w_{i1}, w_{i2}, \cdots, w_{i,m+1}),$$

where

$$w_{ij} = \begin{cases} 1 & \text{if } t^*_{j-1} < t_i \leq t^*_j, \\ 0 & \text{otherwise.} \end{cases}$$

In other words,

$$w_{ij} = \begin{cases} 1 & \text{if } t_i \in I_j, \\ 0 & \text{otherwise,} \end{cases}$$

where, $I_j = (t^*_{j-1}, t^*_j]$, $j = 1, 2, \cdots, m+1$, and from Lawless (2003) we know

$$\mathbf{w}_i \sim \text{Multinominal } (n = 1; P_1, P_2, \cdots, P_{m+1}),$$

where $P_j = P(T \in I_j) = P\big(t^*_{j-1} < T \leq t^*_j\big)$.

For the interval censored data, we only observe the frequencies among $I_1 = (t_0^*, t_1^*], I_2 = (t_1^*, t_2^*], \cdots, I_k = (t_{k-1}^*, t_k^*], \cdots, I_m = (t_{m-1}^*, t_m^*], I_{m+1} = (t_m^*, \infty)$ in which the exact failure time falls.

The data available are denoted in the vector form

$$\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_n.$$

Since these random vectors are independent, the exact likelihood function for $\boldsymbol{\theta}$ conditional on $\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_n$ will be

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^{n} f(\mathbf{w}_i; \boldsymbol{\theta}) = \prod_{i=1}^{n}\prod_{j=1}^{m+1} P_j^{w_{ij}} \\ &= \prod_{i=1}^{n}\left(\prod_{j=1}^{m}\left\{\int_{t_{j-1}^*}^{t_j^*} f(t;\boldsymbol{\theta})\mathrm{d}t\right\}^{w_{ij}}\right)\left\{S\left(t_m^*;\boldsymbol{\theta}\right)\right\}^{w_{i,m+1}} \\ &= \prod_{i=1}^{n}\left(\prod_{j=1}^{m}\left\{F\left(t_j^*;\boldsymbol{\theta}\right) - F\left(t_{j-1}^*;\boldsymbol{\theta}\right)\right\}^{w_{ij}}\right)\left\{S\left(t_m^*;\boldsymbol{\theta}\right)\right\}^{w_{i,m+1}} \\ &= \prod_{i=1}^{n}\left(\prod_{j=1}^{m}\left\{S\left(t_{j-1}^*;\boldsymbol{\theta}\right) - S\left(t_j^*;\boldsymbol{\theta}\right)\right\}^{w_{ij}}\right)\left\{S\left(t_m^*;\boldsymbol{\theta}\right)\right\}^{w_{i,m+1}}, \end{aligned} \qquad (3.1.2)$$

where, t_m^* is the RC time.

3.1.1 Another Representation of the IC Likelihood

Consider another representation of the exact likelihood. By the calculus Mean Value Theorem, for every $\left(t_{j-1}^*, t_j^*\right)$ there must be a point $\tilde{t}_j$ $\left(t_{j-1}^* < \tilde{t}_j < t_j^*\right)$ such that

$$\int_{t_{j-1}^*}^{t_j^*} f(t;\boldsymbol{\theta})\mathrm{d}t = f\left(\tilde{t}_j;\boldsymbol{\theta}\right)\left(t_j^* - t_{j-1}^*\right),$$

then

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^{n}\left(\prod_{j=1}^{m}\left\{f\big(\tilde{t}_j;\boldsymbol{\theta}\big)\big(t_j^*-t_{j-1}^*\big)\right\}^{w_{ij}}\right)\left\{S\big(t_m^*;\boldsymbol{\theta}\big)\right\}^{w_{i,m+1}} \\ &= \prod_{i=1}^{n}\left(\prod_{j=1}^{m}\left\{t_j^*-t_{j-1}^*\right\}^{w_{ij}}\prod_{i=1}^{n}\prod_{j=1}^{m}\left\{f\big(\tilde{t}_j;\boldsymbol{\theta}\big)\right\}^{w_{ij}}\right)\left\{S\big(t_m^*;\boldsymbol{\theta}\big)\right\}^{w_{i,m+1}}. \end{aligned}$$

Since $\prod_{i=1}^{n}\prod_{j=1}^{m}\left\{t_j^*-t_{j-1}^*\right\}^{w_{ij}}$ does not depend on the parameter $\boldsymbol{\theta}$, we have

$$L(\boldsymbol{\theta}) \propto \prod_{i=1}^{n}\left(\prod_{j=1}^{m}\left\{f\big(\tilde{t}_j;\boldsymbol{\theta}\big)\right\}^{w_{ij}}\right)\left\{S\big(t_m^*;\boldsymbol{\theta}\big)\right\}^{w_{i,m+1}}. \tag{3.1.3}$$

3.2 The Pseudo-Likelihood

If we take $\tilde{t}_j = t_j^*$ in (3.1.3), alternatively the midpoint of the j-th interval, we obtain the pseudo-likelihood function

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n}\prod_{j=1}^{m}\left\{f\big(t_j^*;\boldsymbol{\theta}\big)\right\}^{w_{ij}}\left\{S\big(t_m^*;\boldsymbol{\theta}\big)\right\}^{w_{i,m+1}} \tag{3.2.4}$$

which is commonly used in practice. In some longitudinal studies the end-point is used as if it were the true time at which the event of interest occurred, while in other studies the mid-point is used. In most studies the interval-censoring is ignored, and the pseudo likelihood is treated as if it is true likelihood.

Throughout this thesis, we denote any association with pseudo likelihood with subscript *mis* and with true likelihood with the subscript *true*. In general, if covariates are available, (3.1.2) and (3.2.4) can be written as follows

$$L_{true}(\boldsymbol{\alpha},\boldsymbol{\beta}) = \prod_{i=1}^{n}\prod_{j=1}^{m}\left\{S\big(t_{j-1}^*;\boldsymbol{\alpha},\boldsymbol{\beta},\mathrm{x}_i\big)-S\big(t_j^*;\boldsymbol{\alpha},\boldsymbol{\beta},\mathrm{x}_i\big)\right\}^{w_{ij}}\left\{S\big(t_m^*;\boldsymbol{\alpha},\boldsymbol{\beta},\mathrm{x}_i\big)\right\}^{w_{i,m+1}}, \tag{3.2.5}$$

$$L_{mis}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=1}^{n} \prod_{j=1}^{m} \left\{ f\left(t_j^*; \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{x}_i\right) \right\}^{w_{ij}} \left\{ S\left(t_m^*; \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{x}_i\right) \right\}^{w_{i,m+1}}, \tag{3.2.6}$$

where $\boldsymbol{\alpha}$ is the unconstrained parameter vector, $\boldsymbol{\beta}$ is the parameter vector corresponding to covariate vectors $\mathbf{x}_i$ for the i-th subject.

3.3 An Exponential Regression Model

Let T be a continuous non-negative random variable denoting the time to an event of interest. In our investigation, we let T follow exponential distribution in a regression model with $\mathbf{x}$ as a $p \times 1$ vector of fixed covariates. The properties of this model are outlined below. The hazard function for T for the i-th individual takes the form

$$\lambda(t_i) = \lambda_0 \exp(\mathbf{x}_i' \boldsymbol{\beta}), \tag{3.3.7}$$

where λ_0 is a constant unknown underlying hazard to be estimated and $\boldsymbol{\beta}$ is $p \times 1$ vector of regression coefficients. Clearly this is a parametric proportional hazard regression model. Let X be the $n \times p$ matrix with (i, j) entry $x_{ij}, j = 1, 2, \cdots, p$, then $\mathbf{x}_i' \boldsymbol{\beta} = x_{i1}\beta_1 + \cdots + x_{ip}\beta_p$. Of course, $\lambda(t_i) > 0$, so a restriction must be placed on λ_0, namely $0 < \lambda_0 < \infty$. We could also consider $\lambda_0 = \mathrm{e}^{\alpha}$ which would guarantee positive hazard for all α and $\boldsymbol{\beta}$. Let us now briefly address the case when $\lambda_0 = \mathrm{e}^{\alpha}$ where α is some unconstrained parameter.

The hazard for the i-th individual becomes

$$\lambda(t_i) = \mathrm{e}^{\alpha + \mathbf{x}_i' \boldsymbol{\beta}}. \tag{3.3.8}$$

Integrating both sides of (3.3.8),

$$\int_0^{t_i} \lambda(u) \mathrm{d}u = e^{\alpha + \mathbf{x}_i' \boldsymbol{\beta}} \int_0^{t_i} \mathrm{d}u. \tag{3.3.9}$$

Multiplying by -1 and exponentiating both sides of (3.3.9),

$$\exp\Big\{-\int_0^{t_i}\lambda(u)\mathrm{d}u\Big\} = \exp\Big\{-\mathrm{e}^{\alpha+\mathbf{x}_i'\beta}\int_0^{t_i}\mathrm{d}u\Big\}. \tag{3.3.10}$$

From (2.1.1)

$$S(t_i) = \exp\Big\{-\int_0^{t_i}\lambda(u)\mathrm{d}u\Big\},$$

where $S(t_i)$ is the survival function of i-th individual. So, (3.3.10) may be written as

$$S(t_i) = \exp\Big\{-\mathrm{e}^{\alpha+\mathbf{x}_i'\beta}t_i\Big\}$$

$$S(t_i) = S_0(t_i)^{\exp(\mathbf{x}_i'\beta)},$$

where

$$S_0(t_i) = \exp\{-\mathrm{e}^{\alpha}t_i\}$$

is the baseline survival function defined at $\mathbf{x}_i' = \mathbf{0}$ for $\forall\ i$.

3.4 Estimate of Maximum Pseudo-Likelihood

Assume that a censored random sample consisting of data $(t^*_{i,k-1}, t^*_{ik}, \delta_i, \mathbf{x}_i), i = 1, 2, \cdots, n$, is available, where $t^*_{i,k-1}, t^*_{ik}$ are the lower and upper points of the k-th censored interval, and k is the generic subscript which means that i-th individual falls in the k-th interval. The survival time t_i is only observed to lie in the interval $(t^*_{i,k-1}, t^*_{ik}]$ according to whether $\delta_i = 1$ or 0, respectively, hence, w_{ij} in Section 3.1 has the same meaning as δ_i when $j = k$, for an event (IC), but $\delta_i = 0$ can handle any right-censored event. In addition, $\mathbf{x}_i$ is $p \times 1$ covariate vector, $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})'$. Let's suppose that the end points of the interval is used as "exact" time to event.

Then we would like to reexpress the pseudo-likelihood (3.2.4) as follows

$$L_{mis}(\boldsymbol{\theta}) = \prod_{i=1}^{n} \Big\{\lambda\big(t_{ik}^{*};\boldsymbol{\theta}\big)S\big(t_{ik}^{*};\boldsymbol{\theta}\big)\Big\}^{\delta_i}\Big\{S\big(t_i^c;\boldsymbol{\theta}\big)\Big\}^{1-\delta_i}, \tag{3.4.11}$$

where t_i^c, $i = 1, 2, \cdots, n$, denotes any right censored time, e.g., $t_i^c \le t_m^*$, and

$$\delta_i = \begin{cases} 1 \text{ for an event,} \\ 0 \text{ for a right censored observation.} \end{cases}$$

In (3.4.11), $\lambda\big(t_{ik}^*\big)S\big(t_{ik}^*\big)$ is the probability of an interval censored observation and $S\big(t_i^c\big)$ is the probability of associated with the right censored observation. Thus, (3.4.11) is a more general form of the pseudo-likelihood. For the exponential distribution with p covariates we may write (3.4.11) as

$$L_{mis}(\alpha,\boldsymbol{\beta}) = \prod_{i=1}^{n} \Big\{\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}\mathrm{e}^{-\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}t_{ik}^*}\Big\}^{\delta_i}\Big\{\mathrm{e}^{-\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}t_i^c}\Big\}^{1-\delta_i}. \tag{3.4.12}$$

Taking the logarithm on (3.4.12), it yields

$$\ell_{mis}(\alpha,\boldsymbol{\beta}) = \sum_{i=1}^{n}\Big\{\delta_i(\alpha+\mathrm{x}_i'\boldsymbol{\beta}) - \delta_i\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}t_{ik}^* - (1-\delta_i)\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}t_i^c\Big\}.$$

For simplicity, we sometime denote $\ell_{mis}(\alpha,\boldsymbol{\beta})$ as $\partial\ell_{mis}$. The first order partial derivative with respect to α is

$$\begin{aligned}\frac{\partial\ell_{mis}}{\partial\alpha} &= \sum_{i=1}^{n}\Big\{\delta_i - \delta_i\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}t_{ik}^* - (1-\delta_i)\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}t_i^c\Big\} \\ &= \sum_{i=1}^{n}\Big\{\delta_i - \mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}\Big[\delta_i t_{ik}^* + (1-\delta_i)t_i^c\Big]\Big\}.\end{aligned} \tag{3.4.13}$$

At the maximum,

$$\sum_{i=1}^{n}\delta_i = \mathrm{e}^{\alpha}\sum_{i=1}^{n}\mathrm{e}^{\mathrm{x}_i'\boldsymbol{\beta}}\Big[\delta_i t_{ik}^* + (1-\delta_i)t_i^c\Big],$$

hence, the explicit analytic form of maximum likelihood estimate of α is

$$\hat{\alpha} = \log\left[\frac{\sum_{i=1}^{n}\delta_i}{\sum_{i=1}^{n} e^{\mathbf{x}_i'\boldsymbol{\beta}}\left[\delta_i t_{ik}^* + (1-\delta_i)t_i^c\right]}\right],$$

where $\delta_i t_{ik}^* + (1-\delta_i)t_i^c$ implies some time, so we may let it be t_i,

$$\hat{\alpha} = \log\left[\frac{\sum_{i=1}^{n}\delta_i}{\sum_{i=1}^{n} e^{\mathbf{x}_i'\boldsymbol{\beta}} t_i}\right].$$

The first order partial derivatives with respect to the component of $\boldsymbol{\beta}$ are

$$\frac{\partial \ell_{mis}}{\partial \beta_j} = \sum_{i=1}^{n}\left\{\delta_i x_{ij} - \delta_i x_{ij} e^{\alpha+\mathbf{x}_i'\boldsymbol{\beta}} t_{ik}^* - (1-\delta_i)x_{ij} e^{\alpha+\mathbf{x}_i'\boldsymbol{\beta}} t_i^c\right\}, \quad j = 1, 2, \cdots, p. \quad (3.4.14)$$

At the maximum,

$$\sum_{i=1}^{n}\delta_i x_{ij} = \sum_{i=1}^{n} x_{ij} e^{\alpha+\mathbf{x}_i'\hat{\boldsymbol{\beta}}}\left[\delta_i t_{ik}^* + (1-\delta_i)t_i^c\right],$$

yielding,

$$\sum_{i=1}^{n}\delta_i x_{ij} = e^{\alpha}\sum_{i=1}^{n} x_{ij} e^{\mathbf{x}_i'\hat{\boldsymbol{\beta}}} t_i. \quad (3.4.15)$$

Substituting from (3.4.13), we reach

$$\sum_{i=1}^{n}\delta_i x_{ij} = \frac{\sum_{i=1}^{n}\delta_i}{\sum_{i=1}^{n} e^{\mathbf{x}_i'\hat{\boldsymbol{\beta}}} t_i}\sum_{i=1}^{n} x_{ij} e^{\mathbf{x}_i'\hat{\boldsymbol{\beta}}} t_i, \qquad j = 1, 2, \cdots, p$$

which is a system of equations for the MLE of β_j, $\hat{\beta}_j$, $j = 1, 2, \cdots, p$.

3.5 Estimate of Maximum True Likelihood

Under the same assumption and notation as in Section 3.4, similarly, we would like to show the true likelihood (3.1.2) as follows

$$L_{true}(\boldsymbol{\theta}) = \prod_{i=1}^{n}\left\{S\left(t_{i,k-1}^*; \boldsymbol{\theta}\right) - S\left(t_{ik}^*; \boldsymbol{\theta}\right)\right\}^{\delta_i}\left\{S\left(t_i^c; \boldsymbol{\theta}\right)\right\}^{1-\delta_i}, \quad (3.5.16)$$

where, $t^*_{i,k-1}$ and t^*_{ik} are the starting and end points of the intervals for the i-th individual, and

$$\delta_i = \begin{cases} 1 \text{ for an event,} \\ 0 \text{ for a right censored observation.} \end{cases}$$

MacKenzie (1999) proposed a form of the true likelihood function which fully accounts for interval censored, as well as for right censored observations like (3.5.16).

Since

$$\begin{aligned} S(t^*_{i,k-1}) - S(t^*_{ik}) &= \mathrm{e}^{-\int_0^{t^*_{i,k-1}} \lambda(u)\mathrm{d}u} - \mathrm{e}^{-\int_0^{t^*_{ik}} \lambda(u)\mathrm{d}u} \\ &= \mathrm{e}^{-\int_0^{t^*_{i,k-1}} \lambda(u)\mathrm{d}u} - \mathrm{e}^{-\int_0^{t^*_{i,k-1}} \lambda(u)\mathrm{d}u}\mathrm{e}^{-\int_{t^*_{i,k-1}}^{t^*_{ik}} \lambda(u)\mathrm{d}u} \\ &= \mathrm{e}^{-\int_0^{t^*_{i,k-1}} \lambda(u)\mathrm{d}u}\left[1 - \mathrm{e}^{-\int_{t^*_{i,k-1}}^{t^*_{ik}} \lambda(u)\mathrm{d}u}\right] \\ &= S\big(t^*_{i,k-1}\big)\Big[1 - S\big(t^*_{i,k-1}, t^*_{ik}\big)\Big], \end{aligned}$$

then (3.5.16) can be re-written as

$$\begin{aligned} L_{true}(\boldsymbol{\theta}) &= \prod_{i=1}^{n} \Big\{ S\big(t^*_{i,k-1}; \boldsymbol{\theta}\big)\Big[1 - S\big(t^*_{i,k-1}, t^*_{ik}; \boldsymbol{\theta}\big)\Big]\Big\}^{\delta_i} \Big\{ S\big(t^c_i; \boldsymbol{\theta}\big)\Big\}^{1-\delta_i} \\ &= \prod_{i=1}^{n} \Big\{ S\big(t^*_{i,k-1}; \boldsymbol{\theta}\big)\Big[1 - \frac{S\big(t^*_{ik}; \boldsymbol{\theta}\big)}{S\big(t^*_{i,k-1}; \boldsymbol{\theta}\big)}\Big]\Big\}^{\delta_i} \Big\{ S\big(t^c_i; \boldsymbol{\theta}\big)\Big\}^{1-\delta_i}. \end{aligned}$$

Hence, the true likelihood for the exponential regression with p covariates can be written as

$$L_{true}(\alpha, \boldsymbol{\beta}) = \prod_{i=1}^{n} \Big\{ \mathrm{e}^{-\mathrm{e}^{\alpha+\mathbf{x}_i'\boldsymbol{\beta}} t^*_{i,k-1}} \Big[1 - \mathrm{e}^{-\mathrm{e}^{\alpha+\mathbf{x}_i'\boldsymbol{\beta}} d_{ik}(t)}\Big]\Big\}^{\delta_i} \Big\{ \mathrm{e}^{-\mathrm{e}^{\alpha+\mathbf{x}_i'\boldsymbol{\beta}} t^c_i} \Big\}^{1-\delta_i},$$

where $d_{ik}(t) = t^*_{ik} - t^*_{i,k-1}$ is the width of the k-th interval.

The log-likelihood is

$$\ell_{true}(\alpha, \boldsymbol{\beta}) = \sum_{i=1}^{n} \Big\{ -\delta_i \mathrm{e}^{\alpha+\mathbf{x}_i'\boldsymbol{\beta}} t^*_{i,k-1} + \delta_i \log_{\mathrm{e}}\Big[1 - \mathrm{e}^{-\mathrm{e}^{\alpha+\mathbf{x}_i'\boldsymbol{\beta}} d_{ik}(t)}\Big] - (1-\delta_i)\mathrm{e}^{\alpha+\mathbf{x}_i'\boldsymbol{\beta}} t^c_i \Big\}.$$

For simplicity, we sometime denote $\ell_{ture}(\alpha, \beta)$ as $\partial\ell_{true}$. The first order partial derivative with respect to α is

$$\frac{\partial\ell_{true}}{\partial\alpha} = \sum_{i=1}^{n}\left\{ -\delta_i e^{\alpha+x_i'\beta}t_{i,k-1}^* + \delta_i e^{\alpha+x_i'\beta}d_{ik}(t)\frac{e^{-e^{\alpha+x_i'\beta}d_{ik}(t)}}{1-e^{-e^{\alpha+x_i'\beta}d_{ik}(t)}} - (1-\delta_i)e^{\alpha+x_i'\beta}t_i^c\right\}. \tag{3.5.17}$$

But the second term of this derivative makes it impossible to find the explicit form for $\hat{\alpha}$, the best we can do is to find an approximate value serving as a starting point in any Newton-Raphson method. So, in the following expansion of conditional odds on surviving in the interval $(t_{i,k-1}^*, t_{ik}^*]$,

$$\frac{e^{-e^{\alpha+x_i'\beta}d_{ik}(t)}}{1-e^{-e^{\alpha+x_i'\beta}d_{ik}(t)}} = \frac{1 - e^{\alpha+x_i'\beta}d_{ik}(t) + \frac{1}{2}e^{2\alpha+2x_i'\beta}d_i^2(t) - \frac{1}{6}e^{3\alpha+3x_i'\beta}d_i^3(t) + - \cdots}{e^{\alpha+x_i'\beta}d_{ik}(t) - \frac{1}{2}e^{2\alpha+2x_i'\beta}d_i^2(t) + \frac{1}{6}e^{3\alpha+3x_i'\beta}d_i^3(t) - + \cdots},$$

we take the first order approximation, i.e.,

$$\frac{e^{-e^{\alpha+x_i'\beta}d_{ik}(t)}}{1-e^{-e^{\alpha+x_i'\beta}d_{ik}(t)}} \approx \frac{1-e^{\alpha+x_i'\beta}d_{ik}(t)}{e^{\alpha+x_i'\beta}d_{ik}(t)}. \tag{3.5.18}$$

Substituting (3.5.18) back into (3.5.17)

$$\frac{\partial\ell_{true}}{\partial\alpha} = \sum_{i=1}^{n}\left\{ -e^{\alpha+x_i'\beta}t_i' + \delta_i e^{\alpha+x_i'\beta}d_{ik}(t)\frac{1-e^{\alpha+x_i'\beta}d_{ik}(t)}{e^{\alpha+x_i'\beta}d_{ik}(t)}\right\}, \tag{3.5.19}$$

where $t_i' = \delta_i t_{i,k-1}^* + (1-\delta_i)t_i^c$ is some time. At maximum, we have

$$\sum_{i=1}^{n} e^{\hat{\alpha}+x_i'\beta}t_i' = \sum_{i=1}^{n}\left\{\delta_i e^{\hat{\alpha}+x_i'\beta}d_{ik}(t)\frac{1-e^{\hat{\alpha}+x_i'\beta}d_{ik}(t)}{e^{\hat{\alpha}+x_i'\beta}d_{ik}(t)}\right\}$$

$$\sum_{i=1}^{n} e^{\hat{\alpha}+x_i'\beta}\left[t_i' + \delta_i d_{ik}(t)\right] = \sum_{i=1}^{n}\delta_i,$$

from which the approximate analytic form of α is

$$\hat{\alpha}_0 = \log\left[\frac{\sum_{i=1}^{n}\delta_i}{\sum_{i=1}^{n}e^{x_i'\beta}t_i}\right],$$

where as before, we let $t_i' + \delta_i d_{ik}(t) = \delta_i t_{ik}^* + (1-\delta_i)t_i^c$ be any time t_i and where $\hat{\alpha}_0$ indicates that we are only dealing with an approximation.

It is interesting to note that this is the same as $\hat{\alpha}$ resulting from the pseudo-likelihood.

The first order derivatives with respect to the component of $\boldsymbol{\beta}$ are

$$\frac{\partial \ell_{true}}{\partial \beta_j} = \sum_{i=1}^{n} \Bigg\{ - \delta_i x_{ij} \mathrm{e}^{\alpha + \mathbf{x}_i'\boldsymbol{\beta}} t_{i,k-1}^* + \delta_i x_{ij} \mathrm{e}^{\alpha+\mathbf{x}_i'\boldsymbol{\beta}} d_{ik}(t) \frac{\mathrm{e}^{-\mathrm{e}^{\alpha+\mathbf{x}_i'\boldsymbol{\beta}} d_{ik}(t)}}{1 - \mathrm{e}^{-\mathrm{e}^{\alpha+\mathbf{x}_i'\boldsymbol{\beta}} d_{ik}(t)}} - (1-\delta_i) x_{ij} \mathrm{e}^{\alpha+\mathbf{x}_i'\boldsymbol{\beta}} t_i^c \Bigg\}. \quad (3.5.20)$$

As with $\hat{\alpha}$, it is not possible to obtain the analytical expression for $\hat{\boldsymbol{\beta}}$, so, once more, we make the assumption (3.5.18), and then at maximum

$$\sum_{i=1}^{n} x_{ij} \mathrm{e}^{\alpha+\mathbf{x}_i'\hat{\boldsymbol{\beta}}} t_i' = \sum_{i=1}^{n} \Bigg\{ \delta_i x_{ij} \mathrm{e}^{\alpha+\mathbf{x}_i'\hat{\boldsymbol{\beta}}} d_{ik}(t) \frac{1 - \mathrm{e}^{\alpha+\mathbf{x}_i'\hat{\boldsymbol{\beta}}} d_{ik}(t)}{\mathrm{e}^{\alpha+\mathbf{x}_i'\hat{\boldsymbol{\beta}}} d_{ik}(t)} \Bigg\}$$

$$\sum_{i=1}^{n} x_{ij} \mathrm{e}^{\alpha+\mathbf{x}_i'\hat{\boldsymbol{\beta}}} t_i' = \sum_{i=1}^{n} \Big\{ \delta_i x_{ij} - \delta_i x_{ij} \mathrm{e}^{\alpha+\mathbf{x}_i'\hat{\boldsymbol{\beta}}} d_{ik}(t) \Big\}$$

$$\sum_{i=1}^{n} \delta_i x_{ij} = \sum_{i=1}^{n} \Big\{ x_{ij} \mathrm{e}^{\alpha+\mathbf{x}_i'\hat{\boldsymbol{\beta}}} \Big[t_i' + \delta_i d_{ik}(t) \Big] \Big\},$$

yielding

$$\sum_{i=1}^{n} \delta_i x_{ij} = \mathrm{e}^{\alpha} \sum_{i=1}^{n} x_{ij} \mathrm{e}^{\mathbf{x}_i'\hat{\boldsymbol{\beta}}} t_i \quad (3.5.21)$$

which is the same as (3.4.15).

3.6 A Connection between the Likelihoods

For the sake of a clearer connection between the two likelihoods, we could write (3.4.15) as

$$\sum_{i=1}^{n} \delta_i x_{ij} - \sum_{i=1}^{n} x_{ij} \mathrm{e}^{\hat{\alpha}+\mathbf{x}_i'\hat{\boldsymbol{\beta}}} t_i = 0,$$

where we have only substituted $\hat{\alpha}$ for α, then

$$\sum_{i=1}^{n} x_{ij}\left[\delta_i - \mathrm{e}^{\hat{\alpha}+\mathbf{x}_i'\hat{\boldsymbol{\beta}}} t_i\right] = 0,$$

$$\sum_{i=1}^{n} x_{ij}\left[\delta_i - t_i\hat{\lambda}(t_i)\right] = 0, \tag{3.6.22}$$

where $\delta_i - t_i\hat{\lambda}(t_i)$ has the structure of a Martingale residual with δ_i being the observed number of desired events, for the i-th individual, in the interval $(0, t_i)$, and $t_i\hat{\lambda}(t_i)$, the estimated integrated hazard, is the Cox-Snell residual (see Cox and Snell (1968)), which we can regard as the stochastic expectation of the number of desired events in the interval $(0, t_i)$ for the i-th individual.

Similarly, for the case of true likelihood, we can write (3.5.21) as

$$\sum_{i=1}^{n} x_{ij}\left[\delta_i - \mathrm{e}^{\hat{\alpha}+\mathbf{x}_i'\hat{\boldsymbol{\beta}}} t_i\right] = 0,$$

i.e.,

$$\sum_{i=1}^{n} x_{ij}\left[\delta_i - t_i\hat{\lambda}(t_i)\right] = 0$$

which is an identical form to (3.6.22).

3.7 Observed Information Matrix: Pseudo-Likelihood Case

Differentiating (3.4.13) with α again

$$\begin{aligned}\frac{\partial^2 \ell_{mis}}{\partial \alpha^2} &= \sum_{i=1}^{n}\left\{ - \mathrm{e}^{\alpha+\mathbf{x}_i'\boldsymbol{\beta}}\left[\delta_i t_{ik}^* + (1-\delta_i)t_i^c\right]\right\} \\ &= -\sum_{i=1}^{n} \mathrm{e}^{\alpha+\mathbf{x}_i'\boldsymbol{\beta}} t_i. \end{aligned} \tag{3.7.23}$$

Differentiating (3.4.14) with the component of $\boldsymbol{\beta}$ again,

$$\frac{\partial^2 \ell_{mis}}{\partial\beta_j\partial\beta_h} = \sum_{i=1}^{n}\Big\{ -\delta_i x_{ij}x_{ih}\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}t_{ik}^{*} - (1-\delta_i)x_{ij}x_{ih}\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}t_i^{c}\Big\}$$

$$= -\sum_{i=1}^{n} x_{ij}x_{ih}\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}t_i. \tag{3.7.24}$$

Now, from either (3.4.13) or (3.4.14), the mixed second partial derivatives are observed

$$\frac{\partial^2 \ell_{mis}}{\partial\alpha\partial\beta_j} = \sum_{i=1}^{n}\Big\{ -\delta_i x_{ij}\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}t_{ik}^{*} - (1-\delta_i)x_{ij}\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}t_i^{c}\Big\}$$

$$= -\sum_{i=1}^{n} x_{ij}\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}t_i. \tag{3.7.25}$$

The observed information matrix is, in the form

$$I^{[o]}(\alpha,\boldsymbol{\beta}) = \begin{bmatrix} -\partial^2\ell/\partial\alpha^2 & -\partial^2\ell/\partial\alpha\partial\boldsymbol{\beta} \\ -\partial^2\ell/\partial\alpha\partial\boldsymbol{\beta}' & -\partial^2\ell\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' \end{bmatrix}.$$

From (3.7.23), (3.7.24) and (3.7.25), the observed information matrix evaluated at $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$ under pseudo-likelihood has the form

$$I^{[o]}_{mis}(\hat{\alpha},\hat{\boldsymbol{\beta}}) = \begin{bmatrix} \sum_{i=1}^{n}\mathrm{e}^{\hat{\alpha}+\mathrm{x}_i'\hat{\boldsymbol{\beta}}}t_i & \sum_{i=1}^{n}\mathrm{x}_i'\mathrm{e}^{\hat{\alpha}+\mathrm{x}_i'\hat{\boldsymbol{\beta}}}t_i \\ \sum_{i=1}^{n}\mathrm{x}_i\mathrm{e}^{\hat{\alpha}+\mathrm{x}_i'\hat{\boldsymbol{\beta}}}t_i & \sum_{i=1}^{n}\mathrm{x}_i\mathrm{x}_i'\mathrm{e}^{\hat{\alpha}+\mathrm{x}_i'\hat{\boldsymbol{\beta}}}t_i \end{bmatrix}. \tag{3.7.26}$$

3.8 Observed Information Matrix: True Likelihood Case

Let us first find the exact observed information matrix.

For simplicity, we write the first order partial derivative with respect to α, (3.5.17), as

$$\frac{\partial\ell_{true}}{\partial\alpha} = \sum_{i=1}^{n}\Big\{ -\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}t_i' + \delta_i\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}d_{ik}(t)\frac{R_i}{1-R_i}\Big\},$$

where $R_i = \mathrm{e}^{-\mathrm{e}^{\alpha+\mathrm{x}_i'\beta} d_{ik}(t)}$; differentiating with respect to α again, we have

$$\frac{\partial^2 \ell_{true}}{\partial \alpha^2} = \sum_{i=1}^{n} \left\{ -\mathrm{e}^{\alpha+\mathrm{x}_i'\beta} t_i' + \delta_i \mathrm{e}^{\alpha+\mathrm{x}_i'\beta} d_{ik}(t) \frac{R_i}{1-R_i} - \delta_i \mathrm{e}^{2\alpha+2\mathrm{x}_i'\beta} d_i^2(t) \frac{R_i}{(1-R_i)^2} \right\}$$

$$= \sum_{i=1}^{n} \left\{ -\mathrm{e}^{\alpha+\mathrm{x}_i'\beta} t_i' + \delta_i \mathrm{e}^{\alpha+\mathrm{x}_i'\beta} d_{ik}(t) \frac{R_i}{1-R_i} \left[1 - \frac{\mathrm{e}^{\alpha+\mathrm{x}_i'\beta} d_{ik}(t)}{1-R_i} \right] \right\},$$

from which, upon substituting $\frac{1-\mathrm{e}^{\alpha+\mathrm{x}_i'\beta} d_{ik}(t)}{\mathrm{e}^{\alpha+\mathrm{x}_i'\beta} d_{ik}(t)} \approx \frac{R_i}{1-R_i}$, we have

$$\frac{\partial^2 \ell_{true}}{\partial \alpha^2} = -\sum_{i=1}^{n} \mathrm{e}^{\alpha+\mathrm{x}_i'\beta} t_i' \tag{3.8.27}$$

which clearly has a similar form to (3.7.23), except for the difference between the combinations of times, i.e., here we have t_i' and in (3.7.23) we had t_i.

Writing the first order partial derivatives with respect to the component of $\boldsymbol{\beta}$, (3.5.20), as

$$\frac{\partial \ell_{true}}{\partial \beta_j} = \sum_{i=1}^{n} \left\{ -x_{ij} \mathrm{e}^{\alpha+\mathrm{x}_i'\beta} t_i' + \delta_i x_{ij} \mathrm{e}^{\alpha+\mathrm{x}_i'\beta} d_{ik}(t) \frac{R_i}{1-R_i} \right\}$$

and differentiating with respect to the component of $\boldsymbol{\beta}$ again, we get

$$\frac{\partial^2 \ell_{true}}{\partial \beta_j \partial \beta_h} = \sum_{i=1}^{n} \left\{ -x_{ij} x_{ih} \mathrm{e}^{\alpha+\mathrm{x}_i'\beta} t_i' + \frac{\delta_i x_{ij} x_{ih} \mathrm{e}^{\alpha+\mathrm{x}_i'\beta} d_{ik}(t) R_i}{1-R_i} - \frac{\delta_i x_{ij} x_{ih} \mathrm{e}^{2\alpha+2x_i'\beta} d_i^2(t) R_i}{(1-R_i)^2} \right\}$$

$$= \sum_{i=1}^{n} \left\{ -x_{ij} x_{ih} \mathrm{e}^{\alpha+\mathrm{x}_i'\beta} t_i' + \delta_i x_{ij} x_{ih} \mathrm{e}^{\alpha+\mathrm{x}_i'\beta} d_{ik}(t) \frac{R_i}{1-R_i} \left[1 - \frac{\mathrm{e}^{\alpha+\mathrm{x}_i'\beta} d_{ik}(t)}{1-R_i} \right] \right\}$$

and after the first order approximation of R_i, we have

$$\frac{\partial^2 \ell_{true}}{\partial \beta_j \partial \beta_h} = \sum_{i=1}^{n} \left\{ -x_{ij} x_{ih} \mathrm{e}^{\alpha+\mathrm{x}_i'\beta} t_i' + \delta_i x_{ij} x_{ih} \left[1 - \mathrm{e}^{\alpha+\mathrm{x}_i'\beta} d_{ik}(t) \right] \left[1 - \frac{\mathrm{e}^{\alpha+\mathrm{x}_i'\beta} d_{ik}(t)}{\mathrm{e}^{\alpha+\mathrm{x}_i'\beta} d_{ik}(t)} \right] \right\}$$

$$= -\sum_{i=1}^{n} x_{ij} x_{ih} \mathrm{e}^{\alpha+\mathrm{x}_i'\beta} t_i', \tag{3.8.28}$$

except for t_i', this is of identical structure as in the likelihood case (3.7.24).

Finally, the mixed second partial derivatives are

$$\frac{\partial^2 \ell_{truc}}{\partial\alpha\partial\beta_j} = \sum_{i=1}^{n}\left\{ - x_{ij}\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}t_i' + \delta_i x_{ij}\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}d_{ik}(t)\frac{R_i}{1-R_i}\left[1 - \frac{\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}d_{ik}(t)}{1-R_i}\right]\right\}$$

and, again, we impose the approximation of R_i to get:

$$\frac{\partial^2 \ell_{true}}{\partial\alpha\partial\beta_j} \approx -\sum_{i=1}^{n} x_{ij}\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}t_i' \tag{3.8.29}$$

which is the same as (3.7.25) except for t_i'.

From (3.8.27), (3.8.28) and (3.8.29), the observed information matrix evaluated at $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$ under true likelihood has the form:

$$I_{true}^{[o]}(\hat{\alpha},\hat{\boldsymbol{\beta}}) = \begin{bmatrix} \sum_{i=1}^{n} \mathrm{e}^{\hat{\alpha}+\mathrm{x}_i'\hat{\boldsymbol{\beta}}}t_i' & \sum_{i=1}^{n} \mathrm{x}_i'\mathrm{e}^{\hat{\alpha}+\mathrm{x}_i'\hat{\boldsymbol{\beta}}}t_i' \\ \sum_{i=1}^{n} \mathrm{x}_i\mathrm{e}^{\hat{\alpha}+\mathrm{x}_i'\hat{\boldsymbol{\beta}}}t_i' & \sum_{i=1}^{n} \mathrm{x}_i\mathrm{x}_i'\mathrm{e}^{\hat{\alpha}+\mathrm{x}_i'\hat{\boldsymbol{\beta}}}t_i' \end{bmatrix}. \tag{3.8.30}$$

Let us now find the approximate observed information matrix. Differentiating (3.5.19) with respect to α again, we have

$$\begin{aligned}\frac{\partial^2 \ell_{true}}{\partial\alpha^2} &= \sum_{i=1}^{n}\left\{ - \mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}\left[\delta_i t_{ik}^* + (1-\delta_i)t_i^c\right]\right\} \\ &= -\sum_{i=1}^{n} \mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}t_i \end{aligned} \tag{3.8.31}$$

which is the same as (3.7.23).

After the implementation of approximation (3.5.18) to (3.5.20), differentiating with respect to the component of $\boldsymbol{\beta}$ again, we have

$$\begin{aligned}\frac{\partial^2 \ell_{true}}{\partial\beta_j\partial\beta_h} &\approx \sum_{i=1}^{n}\left\{ - x_{ij}x_{ih}\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}\left[\delta_i t_{i,k-1}^* + (1-\delta_i)t_i^c\right] - \delta_i x_{ij}x_{ih}\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}d_{ik}(t)\right\} \\ &= -\sum_{i=1}^{n}\left\{x_{ij}x_{ih}\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}\left[t_i' + \delta_i d_{ik}(t)\right]\right\} \\ &= -\sum_{i=1}^{n} x_{ij}x_{ih}\mathrm{e}^{\alpha+\mathrm{x}_i'\boldsymbol{\beta}}t_i \end{aligned} \tag{3.8.32}$$

which is the same as (3.7.24).

Differentiating (3.5.19) with respect to the component of $\boldsymbol{\beta}$, we obtain the mixed second partial derivatives:

$$\begin{aligned}\frac{\partial^2 \ell_{true}}{\partial\alpha\partial\beta_j} &= -\sum_{i=1}^{n}\left\{x_{ij}\mathrm{e}^{\alpha+\mathbf{x}_i'\boldsymbol{\beta}}\left[t_i' + \delta_i d_{ik}(t)\right]\right\} \\ &= -\sum_{i=1}^{n} x_{ij}\mathrm{e}^{\alpha+\mathbf{x}_i'\boldsymbol{\beta}}t_i. \end{aligned} \tag{3.8.33}$$

From (3.8.31), (3.8.32) and (3.8.33), the observed information matrix evaluated at $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$ under approximated true likelihood has the form

$$\tilde{I}_{true}^{[o]}(\hat{\alpha},\hat{\boldsymbol{\beta}}) = \begin{bmatrix} \sum_{i=1}^{n}\mathrm{e}^{\hat{\alpha}+\mathbf{x}_i'\hat{\boldsymbol{\beta}}}t_i & \sum_{i=1}^{n}\mathbf{x}_i'\mathrm{e}^{\hat{\alpha}+\mathbf{x}_i'\hat{\boldsymbol{\beta}}}t_i \\ \sum_{i=1}^{n}\mathbf{x}_i\mathrm{e}^{\hat{\alpha}+\mathbf{x}_i'\hat{\boldsymbol{\beta}}}t_i & \sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\mathrm{e}^{\hat{\alpha}+\mathbf{x}_i'\hat{\boldsymbol{\beta}}}t_i \end{bmatrix}$$

and it does not come as a surprise that this is identical to $I_{mis}^{[o]}(\hat{\alpha},\hat{\boldsymbol{\beta}})$.

The results obtained in these sections are indeed remarkable. With the true likelihood, we cannot find the exact analytical expressions, however, their approximate expressions are the same as the exact analytical expressions arising from the pseudo-likelihood. Accordingly, the MLEs from the pseudo-likelihood could be used as starting values for the numerical estimation procedures required to estimate the the MLEs in the true likelihood.

Chapter 4

Precision of the Estimators

MacKenzie (1999) and Blagojevic (2002) who first focused on precision, found that in the exponential distribution, the MLE of model parameter λ, from the pseudo-likelihood (mis-specified likelihood), was artificially precise, i.e.,

$$V_{mis}(\hat{\lambda}) < V_{true}(\hat{\lambda}),$$

where, $V_{mis}(\hat{\lambda})$ is the variance of $\hat{\lambda}$ from the pseudo-likelihood, and $V_{true}(\hat{\lambda})$ is the asymptotic (i.e., first order) variance of $\hat{\lambda}$ obtained from the true likelihood, respectively. This result is referred to as the GB conjecture. The Principle of Induction will be used in Section 4.2 to show whether this conjecture is true or not for an exponential regression model. Due to the identical MLEs for the parameters from both likelihoods, we will set the same values for $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$ in the structure of variances.

4.1 Estimation of Covariance Matrix

The approximate variance-covariance matrix $\text{cov}(\hat{\alpha}, \hat{\boldsymbol{\beta}})$ is obtained by inverting the observed information matrix $I^{[o]}(\hat{\alpha}, \hat{\boldsymbol{\beta}})$, i.e.,

$$\text{cov}(\hat{\alpha}, \hat{\boldsymbol{\beta}}) \simeq \left[I^{[o]}(\hat{\alpha}, \hat{\boldsymbol{\beta}})\right]^{-1} = \frac{1}{\det\left[I^{[o]}(\hat{\alpha}, \hat{\boldsymbol{\beta}})\right]} \text{adj}\left[I^{[o]}(\hat{\alpha}, \hat{\boldsymbol{\beta}})\right], \qquad (4.1.1)$$

where, det and adj denote the determinant and adjugate matrix, respectively.

4.1.1 No Covariates

When $p = 0$, i.e., no covariate in the exponential regression model, from (3.7.26) we have

$$I^{[o]}_{mis}(\hat{\alpha}) = \sum_{i=1}^{n} \text{e}^{\hat{\alpha}} t_i,$$

then

$$V_{mis}(\hat{\alpha}) = \frac{1}{\sum_{i=1}^{n} \text{e}^{\hat{\alpha}} t_i} \qquad (4.1.2)$$

and from (3.8.30) we have

$$I^{[o]}_{true}(\hat{\alpha}) = \sum_{i=1}^{n} \text{e}^{\hat{\alpha}} t'_i,$$

then

$$V_{true}(\hat{\alpha}) = \frac{1}{\sum_{i=1}^{n} \text{e}^{\hat{\alpha}} t'_i} \,. \qquad (4.1.3)$$

4.1.2 One Covariate

When $p = 1$, i.e., only one covariate in the exponential regression model, the (3.7.26) can be written as

$$I^{[o]}_{mis}(\hat{\alpha}, \hat{\beta}) = \begin{bmatrix} \sum_{i=1}^{n} E_{i1} t_i & \sum_{i=1}^{n} x_i E_{i1} t_i \\ \sum_{i=1}^{n} x_i E_{i1} t_i & \sum_{i=1}^{n} x_i^2 E_{i1} t_i \end{bmatrix}$$

and from (3.8.30) we have

$$I_{true}^{[o]}(\hat{\alpha},\hat{\beta}) = \begin{bmatrix} \sum_{i=1}^{n} E_{i1}t_i' & \sum_{i=1}^{n} x_iE_{i1}t_i' \\ \sum_{i=1}^{n} x_iE_{i1}t_i' & \sum_{i=1}^{n} x_i^2E_{i1}t_i' \end{bmatrix},$$

where $E_{i1} = e^{\hat{\alpha}+x_i\hat{\beta}}$ is the estimated hazard in this case. Let

$$A_{11} = \sum_{i=1}^{n} E_{i1}t_i,\ A_{12} = A_{21} = \sum_{i=1}^{n} x_iE_{i1}t_i,\ A_{22} = \sum_{i=1}^{n} x_i^2E_{i1}t_i,$$

$$A_{11}' = \sum_{i=1}^{n} E_{i1}t_i',\ A_{12}' = A_{21}' = \sum_{i=1}^{n} x_iE_{i1}t_i',\ A_{22}' = \sum_{i=1}^{n} x_i^2E_{i1}t_i',$$

we get

$$\det\left[I_{mis}^{[o]}(\hat{\alpha},\hat{\beta})\right] = A_{11}A_{22} - A_{12}^2,$$

$$\operatorname{adj}\left[I_{mis}^{[o]}(\hat{\alpha},\hat{\beta})\right] = \begin{bmatrix} A_{22} & -A_{12} \\ -A_{12} & A_{11} \end{bmatrix}$$

and

$$\det\left[I_{true}^{[o]}(\hat{\alpha},\hat{\beta})\right] = A_{11}'A_{22}' - A_{12}'^2,$$

$$\operatorname{adj}\left[I_{true}^{[o]}(\hat{\alpha},\hat{\beta})\right] = \begin{bmatrix} A_{22}' & -A_{12}' \\ -A_{12}' & A_{11}' \end{bmatrix},$$

then the inverse information matrix based on pseudo-likelihood and true likelihood are as follows,

$$\text{cov}_{mis}(\hat{\alpha},\hat{\beta}) = \begin{bmatrix} \frac{A_{22}}{A_{11}A_{22}-A_{12}^2} & \frac{-A_{12}}{A_{11}A_{22}-A_{12}^2} \\ \frac{-A_{12}}{A_{11}A_{22}-A_{12}^2} & \frac{A_{11}}{A_{11}A_{22}-A_{12}^2} \end{bmatrix} \tag{4.1.4}$$

and

$$\text{cov}_{true}(\hat{\alpha},\hat{\beta}) = \begin{bmatrix} \frac{A_{22}'}{A_{11}'A_{22}'-A_{12}'^2} & \frac{-A_{12}'}{A_{11}'A_{22}'-A_{12}'^2} \\ \frac{-A_{12}'}{A_{11}'A_{22}'-A_{12}'^2} & \frac{A_{11}'}{A_{11}'A_{22}'-A_{12}'^2} \end{bmatrix}. \tag{4.1.5}$$

4.1.3 Two or more Covariates

When $p = 2$, i.e., $\mathbf{x}_i'\boldsymbol{\beta} = x_{i1}\beta_1 + x_{i2}\beta_2$, the (3.7.26) and (3.8.30) can be written as:

$$I_{mis}^{[o]}(\hat{\alpha}, \hat{\boldsymbol{\beta}}) = \begin{bmatrix} \sum_{i=1}^n E_{i2}t_i & \sum_{i=1}^n x_{i1}E_{i2}t_i & \sum_{i=1}^n x_{i2}E_{i2}t_i \\ \sum_{i=1}^n x_{i1}E_{i2}t_i & \sum_{i=1}^n x_{i1}^2E_{i2}t_i & \sum_{i=1}^n x_{i1}x_{i2}E_{i2}t_i \\ \sum_{i=1}^n x_{i2}E_{i2}t_i & \sum_{i=1}^n x_{i1}x_{i2}E_{i2}t_i & \sum_{i=1}^n x_{i2}^2E_{i2}t_i \end{bmatrix},$$

$$I_{true}^{[o]}(\hat{\alpha}, \hat{\boldsymbol{\beta}}) = \begin{bmatrix} \sum_{i=1}^n E_{i2}t_i' & \sum_{i=1}^n x_{i1}E_{i2}t_i' & \sum_{i=1}^n x_{i2}E_{i2}t_i' \\ \sum_{i=1}^n x_{i1}E_{i2}t_i' & \sum_{i=1}^n x_{i1}^2E_{i2}t_i' & \sum_{i=1}^n x_{i1}x_{i2}E_{i2}t_i' \\ \sum_{i=1}^n x_{i2}E_{i2}t_i' & \sum_{i=1}^n x_{i1}x_{i2}E_{i2}t_i' & \sum_{i=1}^n x_{i2}^2E_{i2}t_i' \end{bmatrix},$$

where $E_{i2} = \mathrm{e}^{\hat{\alpha}+x_{i1}\hat{\beta}_1+x_{i2}\hat{\beta}_2}$ is the estimated hazard in this case. Let

$$A_{11} = \sum_{i=1}^n E_{i2}t_i,\ A_{12} = A_{21} = \sum_{i=1}^n x_{i1}E_{i2}t_i,\ A_{13} = A_{31} = \sum_{i=1}^n x_{i2}E_{i2}t_i,$$

$$A_{22} = \sum_{i=1}^n x_{i1}^2E_{i2}t_i,\ A_{23} = A_{32} = \sum_{i=1}^n x_{i1}x_{i2}E_{i2}t_i,\ A_{33} = \sum_{i=1}^n x_{i2}^2E_{i2}t_i,$$

$$A_{11}' = \sum_{i=1}^n E_{i2}t_i',\ A_{12}' = A_{21}' = \sum_{i=1}^n x_{i1}E_{i2}t_i',\ A_{13}' = A_{31}' = \sum_{i=1}^n x_{i2}E_{i2}t_i',$$

$$A_{22}' = \sum_{i=1}^n x_{i1}^2E_{i2}t_i',\ A_{23}' = A_{32} = \sum_{i=1}^n x_{i1}x_{i2}E_{i2}t_i',\ A_{33}' = \sum_{i=1}^n x_{i2}^2E_{i2}t_i',$$

we have

$$\det[I_{mis}^{[o]}(\hat{\alpha}, \hat{\boldsymbol{\beta}})] = A_{11}A_{22}A_{33} + 2A_{12}A_{13}A_{23} - A_{22}A_{13}^2 - A_{33}A_{12}^2 - A_{11}A_{23}^2, \quad (4.1.6)$$

$$\operatorname{adj}[I_{mis}^{[o]}(\hat{\alpha}, \hat{\boldsymbol{\beta}})] = \begin{bmatrix} A_{22}A_{33} - A_{23}^2 & -(A_{12}A_{33} - A_{13}A_{23}) & A_{12}A_{23} - A_{13}A_{22} \\ -(A_{12}A_{33} - A_{13}A_{23}) & A_{11}A_{33} - A_{13}^2 & -(A_{11}A_{23} - A_{12}A_{13}) \\ A_{12}A_{23} - A_{13}A_{22} & -(A_{11}A_{23} - A_{12}A_{13}) & A_{11}A_{22} - A_{12}^2 \end{bmatrix} \quad (4.1.7)$$

and

$$\det\left[I^{[o]}_{true}(\hat{\alpha},\hat{\boldsymbol{\beta}})\right] = A'_{11}A'_{22}A'_{33} + 2A'_{12}A'_{13}A'_{23} - A'_{22}A'^{2}_{13} - A'_{33}A_{12}'2 - A'_{11}A'^{2}_{23}, \quad (4.1.8)$$

$$\operatorname{adj}\left[I^{[o]}_{true}(\hat{\alpha},\hat{\boldsymbol{\beta}})\right] = \begin{bmatrix} A'_{22}A'_{33} - A'^{2}_{23} & -(A'_{12}A'_{33} - A'_{13}A'_{23}) & A'_{12}A'_{23} - A'_{13}A'_{22} \\ -(A'_{12}A'_{33} - A'_{13}A'_{23}) & A'_{11}A'_{33} - A'^{2}_{13} & -(A'_{11}A'_{23} - A'_{12}A'_{13}) \\ A'_{12}A'_{23} - A'_{13}A'_{22} & -(A'_{11}A'_{23} - A'_{12}A'_{13}) & A'_{11}A'_{22} - A'^{2}_{12} \end{bmatrix}. \quad (4.1.9)$$

The corresponding variance-covariance matrixes can be obtained easily from (4.1.6) to (4.1.9). We have

$$V_{mis}(\hat{\alpha}) = \frac{A_{22}A_{33} - A^2_{23}}{A_{11}A_{22}A_{33} + 2A_{12}A_{13}A_{23} - A_{22}A^2_{13} - A_{33}A^2_{12} - A_{11}A^2_{23}}, \quad (4.1.10)$$

$$V_{mis}(\hat{\beta}_1) = \frac{A_{11}A_{33} - A^2_{13}}{A_{11}A_{22}A_{33} + 2A_{12}A_{13}A_{23} - A_{22}A^2_{13} - A_{33}A^2_{12} - A_{11}A^2_{23}}, \quad (4.1.11)$$

$$V_{mis}(\hat{\beta}_2) = \frac{A_{11}A_{22} - A^2_{12}}{A_{11}A_{22}A_{33} + 2A_{12}A_{13}A_{23} - A_{22}A^2_{13} - A_{33}A^2_{12} - A_{11}A^2_{23}}, \quad (4.1.12)$$

$$V_{true}(\hat{\alpha}) = \frac{A'_{22}A'_{33} - A'^{2}_{23}}{A'_{11}A'_{22}A'_{33} + 2A'_{12}A'_{13}A'_{23} - A'_{22}A'^{2}_{13} - A'_{33}A_{12}'2 - A'_{11}A'^{2}_{23}}, \quad (4.1.13)$$

$$V_{true}(\hat{\beta}_1) = \frac{A'_{11}A'_{33} - A'^{2}_{13}}{A'_{11}A'_{22}A'_{33} + 2A'_{12}A'_{13}A'_{23} - A'_{22}A'^{2}_{13} - A'_{33}A_{12}'2 - A'_{11}A'^{2}_{23}}, \quad (4.1.14)$$

$$V_{true}(\hat{\beta}_2) = \frac{A'_{11}A'_{22} - A'^{2}_{12}}{A'_{11}A'_{22}A'_{33} + 2A'_{12}A'_{13}A'_{23} - A'_{22}A'^{2}_{13} - A'_{33}A_{12}'2 - A'_{11}A'^{2}_{23}}. \quad (4.1.15)$$

When we have p covariates in the model, i.e., $\mathbf{x}'_i\boldsymbol{\beta} = x_{i1}\beta_1 + \cdots + x_{ip}\beta_p$, the (3.7.26) and (3.8.30) can be written as

$$I^{[o]}_{mis}(\hat{\alpha},\hat{\boldsymbol{\beta}}) = \begin{bmatrix} \sum_{i=1}^n E_{ip}t_i & \sum_{i=1}^n x_{i1}E_{ip}t_i & \cdots & \sum_{i=1}^n x_{ip}E_{ip}t_i \\ \sum_{i=1}^n x_{i1}E_{ip}t_i & \sum_{i=1}^n x^2_{i1}E_{ip}t_i & \cdots & \sum_{i=1}^n x_{i1}x_{ip}E_{ip}t_i \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^n x_{ip}E_{ip}t_i & \sum_{i=1}^n x_{i1}x_{ip}E_{ip}t_i & \cdots & \sum_{i=1}^n x^2_{ip}E_{ip}t_i \end{bmatrix} \quad (4.1.16)$$

$$I^{[o]}_{true}(\hat{\alpha},\hat{\boldsymbol{\beta}}) = \begin{bmatrix} \sum_{i=1}^{n} E_{ip}t'_i & \sum_{i=1}^{n} x_{i1}E_{ip}t'_i & \cdots & \sum_{i=1}^{n} x_{ip}E_{ip}t'_i \\ \sum_{i=1}^{n} x_{i1}E_{ip}t'_i & \sum_{i=1}^{n} x_{i1}^2 E_{ip}t'i & \cdots & \sum_{i=1}^{n} x_{i1}x_{ip}E_{ip}t'_i \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} x_{ip}E_{ip}t'_i & \sum_{i=1}^{n} x_{i1}x_{ip}E_{ip}t'_i & \cdots & \sum_{i=1}^{n} x_{ip}^2 E_{ip}t'_i \end{bmatrix}, \quad (4.1.17)$$

where $E_{ip} = e^{\hat{\alpha}+x_{i1}\hat{\beta}_1+\cdots+x_{ip}\hat{\beta}_p}$ is the estimated hazard in this case.

The corresponding estimated asymptotic covariance matrix $\text{cov}_{mis}(\hat{\alpha},\hat{\boldsymbol{\beta}})$ and $\text{cov}_{true}(\hat{\alpha},\hat{\boldsymbol{\beta}})$ can be obtained from (4.1.16) and (4.1.17) respectively.

4.2 Comparative Inference

As mentioned above, the comparative inference envisioned is to compare the variances obtained from the pseudo-likelihood with that from the true likelihood for the interval censoring exponential regression model. Since the non-linear conditional odds on surviving in the interval exists in the score functions (3.5.17) and (3.5.20) in the true likelihood case, we could not obtain the explicit estimates for the parameters α and $\boldsymbol{\beta}$, however, the first order approximate estimates are identical to that from pseudo-likelihood (see Chapter 3), then we set the same estimates of parameters at $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$ when we do the analytical comparison next.

Let us start with no covariate in the exponential model. From (4.1.2) and (4.1.3), we have

$$\frac{V_{mis}(\hat{\alpha})}{V_{true}(\hat{\alpha})} = \frac{\sum_{i=1}^{n} e^{\hat{\alpha}}t'_i}{\sum_{i=1}^{n} e^{\hat{\alpha}}t_i} = \frac{\sum_{i=1}^{n} t'_i}{\sum_{i=1}^{n} t_i} < 1,$$

since $t'_i = \delta_i t^*_{i,k-1} + (1-\delta_i)t^c_i < t_i = \delta_i t^*_{ik} + (1-\delta_i)t^c_i$ which shows that the estimator based on pseudo-likelihood under-estimates the true variance $V_{true}(\hat{\alpha})$ when the observed inspection times are analyzed as if they were exact.

Next we think about only one covariate x_i in the exponential model and let $x_i = 0, 1$ be a binary variable, from (4.1.4) and (4.1.5), we have $A_{12} = A_{22}, A'_{12} = A'_{22}$, then

$$\begin{aligned}\frac{V_{mis}(\hat{\alpha})}{V_{true}(\hat{\alpha})} &= \frac{A_{22}}{A_{11}A_{22} - A_{12}^2} \times \frac{A'_{11}A'_{22} - A'^2_{12}}{A'_{22}} \\ &= \frac{A'_{11} - A'_{12}}{A_{11} - A_{12}}\end{aligned} \tag{4.2.18}$$

and

$$\begin{aligned}\frac{V_{mis}(\hat{\beta})}{V_{true}(\hat{\beta})} &= \frac{A_{11}}{A_{11}A_{22} - A_{12}^2} \times \frac{A'_{11}A'_{22} - A'^2_{12}}{A'_{11}} \\ &= \frac{A_{11}(A'_{11}A'_{12} - A'^2_{12})}{A'_{11}(A_{11}A_{12} - A_{12}^2)} \ .\end{aligned} \tag{4.2.19}$$

Denote $i_1 = \{i : x_i = 1\}, i_0 = \{i : x_i = 0\}$, from (4.2.18), we get

$$\begin{aligned}(A_{11} - A_{12}) - (A'_{11} - A'_{12}) &= (A_{11} - A'_{11}) - (A_{12} - A'_{12}) \\ &= \sum_{i=1}^{n} E_{i1}(t_i - t'_i) - \sum_{i=1}^{n} x_i E_{i1}(t_i - t'_i) \\ &= \sum_{i=1}^{n} E_{i1}(t_i - t'_i)(1 - x_i) \\ &= \sum_{i_0} e^{\hat{\alpha}}(t_i - t'_i) > 0,\end{aligned}$$

thus $V_{mis}(\hat{\alpha}) < V_{true}(\hat{\alpha})$.

From (4.2.19), we have

$$
\begin{aligned}
&A'_{11}(A_{11}A_{12}-A_{12}^2)-A_{11}(A'_{11}A'_{12}-A'^2_{12})\\
&=A'_{11}(A_{11}-A_{12})A_{12}-A_{11}(A'_{11}-A'_{12})A'_{12}\\
&=\Big(\sum_{i_1}\mathrm{e}^{\hat\alpha+\hat\beta}t'_i+\sum_{i_0}\mathrm{e}^{\hat\alpha}t'_i\Big)\Big(\sum_{i_1}\mathrm{e}^{\hat\alpha+\hat\beta}t_i+\sum_{i_0}\mathrm{e}^{\hat\alpha}t_i-\sum_{i_1}\mathrm{e}^{\hat\alpha+\hat\beta}t_i\Big)\sum_{i_1}\mathrm{e}^{\hat\alpha+\hat\beta}t_i\\
&-\Big(\sum_{i_1}\mathrm{e}^{\hat\alpha+\hat\beta}t_i+\sum_{i_0}\mathrm{e}^{\hat\alpha}t_i\Big)\Big(\sum_{i_1}\mathrm{e}^{\hat\alpha+\hat\beta}t'_i+\sum_{i_0}\mathrm{e}^{\hat\alpha}t'_i-\sum_{i_1}\mathrm{e}^{\hat\alpha+\hat\beta}t'_i\Big)\sum_{i_1}\mathrm{e}^{\hat\alpha+\hat\beta}t'_i\\
&=\Big(\sum_{i_1}\mathrm{e}^{\hat\alpha+\hat\beta}t'_i+\sum_{i_0}\mathrm{e}^{\hat\alpha}t'_i\Big)\sum_{i_0}\mathrm{e}^{\hat\alpha}t_i\sum_{i_1}\mathrm{e}^{\hat\alpha+\hat\beta}t_i\\
&-\Big(\sum_{i_1}\mathrm{e}^{\hat\alpha+\hat\beta}t_i+\sum_{i_0}\mathrm{e}^{\hat\alpha}t_i\Big)\sum_{i_0}\mathrm{e}^{\hat\alpha}t'_i\sum_{i_1}\mathrm{e}^{\hat\alpha+\hat\beta}t'_i\\
&=\sum_{i_1}\mathrm{e}^{\hat\alpha+\hat\beta}t_i\sum_{i_1}\mathrm{e}^{\hat\alpha+\hat\beta}t'_i\sum_{i_0}\mathrm{e}^{\hat\alpha}(t_i-t'_i)+\sum_{i_0}\mathrm{e}^{\hat\alpha}t_i\sum_{i_0}\mathrm{e}^{\hat\alpha}t'_i\sum_{i_1}\mathrm{e}^{\hat\alpha+\hat\beta}(t_i-t'_i)>0,
\end{aligned}
$$

hence, $V_{mis}(\hat\beta) < V_{true}(\hat\beta)$.

Further more, we consider two categorical covariates x_{i1} and x_{i2} in the model, where

$$x_{i1}=\begin{cases}1 \text{ for the first classification,}\\ 0 \text{ otherwise,}\end{cases}$$

and

$$x_{i2}=\begin{cases}1 \text{ for the second classification,}\\ 0 \text{ otherwise,}\end{cases}$$

in this case, $x_{i1}x_{i2}=0\ \forall\ i=1,2,\cdots,n$, which means that x_{i1} and x_{i2} are orthogonal. Look back to (4.1.10) to (4.1.15), where, $A_{12}=A_{22}, A_{13}=A_{33}, A'_{12}=A'_{22}, A'_{13}=$

$A'_{33}, A_{23} = A'_{23} = 0$, then

$$\frac{V_{mis}(\hat{\alpha})}{V_{true}(\hat{\alpha})} = \frac{A_{22}A_{33}}{A_{11}A_{22}A_{33} - A_{22}A_{13}^2 - A_{33}A_{12}^2} \times \frac{A'_{11}A'_{22}A'_{33} - A'_{22}A'^2_{13} - A'_{33}A'^2_{12}}{A'_{22}A'_{33}}$$
$$= \frac{A'_{11} - A'_{13} - A'_{12}}{A_{11} - A_{13} - A_{12}} \tag{4.2.20}$$

$$\frac{V_{mis}(\hat{\beta}_1)}{V_{true}(\hat{\beta}_1)} = \frac{A_{11}A_{33} - A_{13}^2}{A_{11}A_{22}A_{33} - A_{22}A_{13}^2 - A_{33}A_{12}^2} \times \frac{A'_{11}A'_{22}A'_{33} - A'_{22}A'^2_{13} - A'_{33}A'^2_{12}}{A'_{11}A'_{33} - A'^2_{13}}$$
$$= \frac{(A_{11} - A_{13})(A'_{11}A'_{12} - A'_{12}A'_{13} - A'^2_{12})}{(A'_{11} - A'_{13})(A_{11}A_{12} - A_{12}A_{13} - A_{12}^2)} \tag{4.2.21}$$

$$\frac{V_{mis}(\hat{\beta}_2)}{V_{true}(\hat{\beta}_2)} = \frac{A_{11}A_{22} - A_{12}^2}{A_{11}A_{22}A_{33} - A_{22}A_{13}^2 - A_{33}A_{12}^2} \times \frac{A'_{11}A'_{22}A'_{33} - A'_{22}A'^2_{13} - A'_{33}A'^2_{12}}{A'_{11}A'_{22} - A'^2_{12}}$$
$$= \frac{(A_{11} - A_{12})(A'_{11}A'_{13} - A'_{12}A'_{13} - A'^2_{13})}{(A'_{11} - A'_{12})(A_{11}A_{13} - A_{12}A_{13} - A_{13}^2)} \, . \tag{4.2.22}$$

Denote $i_{10} = \{i : x_{i1} = 1\}, i_{01} = \{i : x_{i2} = 1\}, i_{00} = \{i : x_{i1} = 0 \ \& \ x_{i2} = 0\}$, from (4.2.20), we have

$$(A_{11} - A_{13} - A_{12}) - (A'_{11} - A'_{13} - A'_{12})$$
$$= (A_{11} - A'_{11}) - (A_{12} - A'_{12}) - (A_{13} - A'_{13})$$

$$= \sum_{i=1}^{n} E_{i2}(t_i - t'_i) - \sum_{i=1}^{n} x_{i1}E_{i2}(t_i - t'_i) - \sum_{i=1}^{n} x_{i2}E_{i2}(t_i - t'_i)$$
$$= \sum_{i=1}^{n} \mathrm{e}^{\hat{\alpha}+x_{i1}\hat{\beta}_1+x_{i2}\hat{\beta}_2}(t_i - t'_i)(1 - x_{i1} - x_{i2})$$
$$= \sum_{i_{00}} \mathrm{e}^{\hat{\alpha}}(t_i - t'_i) > 0,$$

thus $V_{mis}(\hat{\alpha}) < V_{true}(\hat{\alpha})$.

From (4.2.21), we have

$$
\begin{aligned}
&(A'_{11} - A'_{13})(A_{11}A_{12} - A_{12}A_{13} - A_{12}^2) - (A_{11} - A_{13})(A'_{11}A'_{12} - A'_{12}A'_{13} - A'^2_{12}) \\
&\quad = (A'_{11} - A'_{13})(A_{11} - A_{12} - A_{13})A_{12} - (A_{11} - A_{13})(A'_{11} - A'_{12} - A'_{13})A'_{12} \\
&\quad = \left(\sum_{i_{10}} e^{\hat{\alpha}+\hat{\beta}_1} t'_i + \sum_{i_{00}} e^{\hat{\alpha}} t'_i\right) \sum_{i_{00}} e^{\hat{\alpha}} t_i \sum_{i_{10}} e^{\hat{\alpha}+\hat{\beta}_1} t_i \\
&\quad\quad - \left(\sum_{i_{10}} e^{\hat{\alpha}+\hat{\beta}_1} t_i + \sum_{i_{00}} e^{\hat{\alpha}} t_i\right) \sum_{i_{00}} e^{\hat{\alpha}} t'_i \sum_{i_{10}} e^{\hat{\alpha}+\hat{\beta}_1} t'_i \\
&\quad = \sum_{i_{10}} e^{\hat{\alpha}+\hat{\beta}_1} t_i \sum_{i_{10}} e^{\hat{\alpha}+\hat{\beta}_1} t'_i \sum_{i_{00}} e^{\hat{\alpha}} (t_i - t'_i) + \sum_{i_{00}} e^{\hat{\alpha}} t_i \sum_{i_{00}} e^{\hat{\alpha}} t'_i \sum_{i_{10}} e^{\hat{\alpha}+\hat{\beta}_1} (t_i - t'_i) > 0,
\end{aligned}
$$

from which, we see, $V_{mis}(\hat{\beta}_1) < V_{true}(\hat{\beta}_1)$.

From (4.2.22), by symmetry, it is easily to show that

$$
\begin{aligned}
&(A'_{11} - A'_{12})(A_{11}A_{13} - A_{12}A_{13} - A_{13}^2) - (A_{11} - A_{12})(A'_{11}A'_{13} - A'_{12}A'_{13} - A'^2_{13}) \\
&\quad = \sum_{i_{01}} e^{\hat{\alpha}+\hat{\beta}_2} t_i \sum_{i_{01}} e^{\hat{\alpha}+\hat{\beta}_2} t'_i \sum_{i_{00}} e^{\hat{\alpha}} (t_i - t'_i) + \sum_{i_{00}} e^{\hat{\alpha}} t_i \sum_{i_{00}} e^{\hat{\alpha}} t'_i \sum_{i_{01}} e^{\hat{\alpha}+\hat{\beta}_2} (t_i - t'_i) > 0,
\end{aligned}
$$

thus $V_{mis}(\hat{\beta}_2) < V_{true}(\hat{\beta}_2)$.

Consider $p = k$ categorical covariates in the model, in this case, we let

$$
x_{ij} = \begin{cases} 1 \text{ for the } j\text{-th classification}, & j = 1, 2, \cdots, k, \\ 0 \text{ otherwise.} \end{cases}
$$

The corresponding variances are

$$
V_{mis}(\hat{\alpha}) = \frac{A_{22}A_{33}\cdots A_{kk}}{A};\ V_{mis}(\hat{\beta}_1) = \frac{A_1}{A},\ \cdots, V_{mis}(\hat{\beta}_j) = \frac{A_j}{A},\ \cdots, V_{mis}(\hat{\beta}_k) = \frac{A_k}{A} \tag{4.2.23}
$$

and

$$
V_{true}(\hat{\alpha}) = \frac{A'_{22}A'_{33}\cdots A'_{kk}}{A'};\ V_{true}(\hat{\beta}_1) = \frac{A'_1}{A'},\ \cdots, V_{true}(\hat{\beta}_j) = \frac{A'_j}{A'},\ \cdots, V_{true}(\hat{\beta}_k) = \frac{A'_k}{A'}, \tag{4.2.24}
$$

where

$$A = A_{11}A_{22}A_{33}\cdots A_{k+1,k+1} - A_{33}A_{44}\cdots A_{k+1,k+1}A_{12}^2 - A_{22}A_{44}\cdots A_{k+1,k+1}A_{13}^2 - \cdots - A_{22}A_{33}\cdots A_{kk}A_{1,k+1}^2$$

$$A_1 = A_{11}A_{33}A_{44}\cdots A_{k+1,k+1} - A_{44}A_{55}\cdots A_{k+1,k+1}A_{13}^2 - A_{33}A_{55}\cdots A_{k+1,k+1}A_{14}^2 - \cdots - A_{33}A_{44}\cdots A_{kk}A_{1,k+1}^2$$

$$\vdots$$

$$\begin{aligned} A_j = {} & A_{11}A_{22}A_{33}\cdots A_{jj}A_{j+2,j+2}\cdots A_{k+1,k+1} \\ & - A_{33}A_{44}\cdots A_{jj}A_{j+2,j+2}\cdots A_{k+1,k+1}A_{12}^2 \\ & - \cdots A_{22}A_{33}\cdots A_{j-1,j-1}A_{j+2,j+2}\cdots A_{k+1,k+1}A_{1j}^2 \\ & - A_{22}A_{33}\cdots A_{jj}A_{j+3,j+3}\cdots A_{k+1,k+1}A_{1,j+2}^2 \\ & - \cdots - A_{22}A_{33}\cdots A_{jj}A_{j+2,j+2}\cdots A_{kk}A_{1,k+1}^2 \end{aligned}$$

$$\vdots$$

$$A_k = A_{11}A_{22}A_{33}\cdots A_{kk} - A_{33}A_{44}\cdots A_{kk}A_{12}^2 - A_{22}A_{44}\cdots A_{kk}A_{13}^2 - \cdots - A_{22}A_{33}\cdots A_{k-1,k-1}A_{1k}^2$$

$$A' = A'_{11}A'_{22}A'_{33}\cdots A'_{k+1,k+1} - A'_{33}A'_{44}\cdots A'_{k+1,k+1}A'^2_{12} - A'_{22}A'_{44}\cdots A'_{k+1,k+1}A'^2_{13} - \cdots - A'_{22}A'_{33}\cdots A'_{kk}A'^2_{1,k+1}$$

$$A'_1 = A'_{11}A'_{33}A'_{44}\cdots A'_{k+1,k+1} - A'_{44}A'_{55}\cdots A'_{k+1,k+1}A'^2_{13}$$
$$- A'_{33}A'_{55}\cdots A'_{k+1,k+1}A'^2_{14} - \cdots - A'_{33}A'_{44}\cdots A'_{kk}A'^2_{1,k+1}$$
$$\vdots$$
$$A'_j = A'_{11}A'_{22}A'_{33}\cdots A'_{jj}A'_{j+2,j+2}\cdots A'_{k+1,k+1}$$
$$- A'_{33}A'_{44}\cdots A'_{jj}A'_{j+2,j+2}\cdots A'_{k+1,k+1}A'^2_{12}$$
$$- \cdots - A'_{22}A'_{33}\cdots A'_{j-1,j-1}A'_{j+2,j+2}\cdots A'_{k+1,k+1}A'^2_{1j}$$
$$- A'_{22}A'_{33}\cdots A'_{jj}A'_{j+3,j+3}\cdots A'_{k+1,k+1}A'^2_{1,j+2}$$
$$- \cdots - A'_{22}A'_{33}\cdots A'_{jj}A'_{j+2,j+2}\cdots A'_{kk}A'^2_{1,k+1}$$
$$\vdots$$
$$A'_k = A'_{11}A'_{22}A'_{33}\cdots A'_{kk} - A'_{33}A'_{44}\cdots A'_{kk}A'^2_{12}$$
$$- A'_{22}A'_{44}\cdots A'_{kk}A'^2_{13} - \cdots - A'_{22}A'_{33}\cdots A'_{k-1,k-1}A'^2_{1k}$$

and

$$A_{11} = \sum_{i=1}^{n} E_{ik}t_i,\ A_{1,j+1} = \sum_{i=1}^{n} x_{ij}E_{ik}t_i = A_{j+1,j+1} = \sum_{i=1}^{n} x_{ij}^2 E_{ik}t_i,$$
$$A'_{11} = \sum_{i=1}^{n} E_{ik}t'_i,\ A'_{1,j+1} = \sum_{i=1}^{n} x_{ij}E_{ik}t'_i = A'_{j+1,j+1} = \sum_{i=1}^{n} x_{ij}^2 E_{ik}t'_i,$$

for $j = 1, 2\cdots, k$. Where, $E_{ik} = \mathrm{e}^{\hat{\alpha}+x_{i1}\beta_1+x_{i2}\beta_2+\cdots+x_{ik}\beta_k}$. These structures of variance in k categorical case were obtained from $p = 1$, $p = 2$ cases, especially from $p = 3$ with the Block partition shown in Section 2.4 used.

Then we have

$$\frac{V_{mis}(\hat{\alpha})}{V_{true}(\hat{\alpha})} = \frac{A_{22}A_{33}\cdots A_{k+1,k+1}}{A} \times \frac{A'}{A'_{22}A'_{33}\cdots A'_{k+1,k+1}}$$
$$= \frac{A'_{11} - A'_{12} - A'_{13} - \cdots - A'_{1,k+1}}{A_{11} - A_{12} - A_{13} - \cdots - A_{1,k+1}}.$$

Denote $i_{10\cdots 0} = \{i : x_{i1} = 1\}, i_{010\cdots 0} = \{i : x_{i2} = 1\}, \cdots, i_{00\cdots 10\cdots 0} = \{i : x_{ij} = 1\}, \cdots, i_{0\cdots 01} = \{i : x_{ik} = 1\}, i_{00\cdots 0} = \{i : x_{i1} = 0, x_{i2} = 0, \cdots, x_{ik} = 0\}$, we have

$$\begin{aligned}
&(A_{11} - A_{12} - A_{13} - \cdots - A_{1,k+1}) - (A'_{11} - A'_{12} - A'_{13} - \cdots - A'_{1,k+1}) \\
&\quad = (A_{11} - A'_{11}) - (A_{12} - A'_{12}) - (A_{13} - A'_{13}) - \cdots - (A_{1,k+1} - A'_{1,k+1}) \\
&\quad = \sum_{i=1}^{n} E_{ik}(t_i - t'_i) - \sum_{i=1}^{n} x_{i1} E_{ik}(t_i - t'_i) \\
&\quad - \sum_{i=1}^{n} x_{i2} E_{ik}(t_i - t'_i) - \cdots - \sum_{i=1}^{n} x_{ik} E_{ik}(t_i - t'_i) \\
&\quad = \sum_{i=1}^{n} \mathrm{e}^{\hat{\alpha} + x_{i1}\hat{\beta}_1 + x_{i2}\hat{\beta}_2 + \cdots + x_{ik}\hat{\beta}_k} (t_i - t'_i)(1 - x_{i1} - x_{i2} - \cdots - x_{ik}) \\
&\quad = \sum_{i_{00\cdots 0}} \mathrm{e}^{\hat{\alpha}} (t_i - t'_i) > 0. \qquad (4.2.25)
\end{aligned}$$

From (4.2.23) and (4.2.24), we have

$$\begin{aligned}
\frac{V_{mis}(\hat{\beta}_j)}{V_{true}(\hat{\beta}_j)} &= \frac{A_j}{A} \times \frac{A'}{A'_j} \\
&= \frac{A_{11} - A_{12} - \cdots - A_{1j} - A_{1,j+2} - \cdots - A_{1,k+1}}{(A_{11} - A_{12} - A_{13} - \cdots - A_{1,k+1}) A_{1j}} \\
&\times \frac{(A'_{11} - A'_{12} - A'_{13} - \cdots - A'_{1,k+1}) A'_{12}}{A'_{11} - A'_{12} - \cdots - A'_{1j} - A'_{1,j+2} - \cdots - A'_{1,k+1}} \\
= &\frac{(A_{11} - A_{12} - \cdots - A_{1j} - A_{1,j+2} - \cdots - A_{1,k+1})(A'_{11} - A'_{12} - A'_{13} - \cdots - A'_{1,k+1}) A'_{1j}}{(A'_{11} - A'_{12} - \cdots - A'_{1j} - A'_{1,j+2} - \cdots - A'_{1,k+1})(A_{11} - A_{12} - A_{13} - \cdots - A_{1,k+1}) A_{1j}},
\end{aligned}$$

from which, we have

$$(A'_{11}-A'_{12}-\cdots-A'_{1j}-A'_{1,j+2}-\cdots-A'_{1,k+1})(A_{11}-A_{12}-A_{13}-\cdots-A_{1,k+1})A_{1j}$$
$$-(A_{11}-A_{12}-\cdots-A_{1j}-A_{1,j+2}-\cdots-A_{1,k+1})(A'_{11}-A'_{12}-A'_{13}-\cdots-A'_{1,k+1})A'_{1j}$$
$$=\Big(\sum_{i_{00\cdots10\cdots0}} e^{\hat{\alpha}+\hat{\beta}_j}t'_i+\sum_{i_{00\cdots0}} e^{\hat{\alpha}}t'_i\Big)\sum_{i_{00\cdots0}} e^{\hat{\alpha}}t_i\sum_{i_{00\cdots10\cdots0}} e^{\hat{\alpha}+\hat{\beta}_j}t_i$$
$$-\Big(\sum_{i_{00\cdots10\cdots0}} e^{\hat{\alpha}+\hat{\beta}_j}t_i+\sum_{i_{00\cdots0}} e^{\hat{\alpha}}t_i\Big)\sum_{i_{00\cdots0}} e^{\hat{\alpha}}t'_i\sum_{i_{00\cdots10\cdots0}} e^{\hat{\alpha}+\hat{\beta}_j}t'_i$$
$$=\sum_{i_{00\cdots10\cdots0}} e^{\hat{\alpha}+\hat{\beta}_j}t_i\sum_{i_{00\cdots10\cdots0}} e^{\hat{\alpha}+\hat{\beta}_j}t'_i\sum_{i_{00\cdots0}} e^{\hat{\alpha}}(t_i-t'_i)+\sum_{i_{00\cdots0}} e^{\hat{\alpha}}t_i\sum_{i_{00\cdots0}} e^{\hat{\alpha}}t'_i\sum_{i_{00\cdots10\cdots0}} e^{\hat{\alpha}+\hat{\beta}_j}(t_i-t'_i)>0,$$
(4.2.26)

for $j=1,2,\cdots,k$. We have demonstrated that $V_{mis}(\hat{\alpha})<V_{true}(\hat{\alpha})$ and $V_{mis}(\hat{\beta}_j)<V_{true}(\hat{\beta}_j), j=1,2,\cdots,p$, hold at $p=0,1,2$. Assume this is true for $p=k$, then consider $p=k+1$ where we have

$$\frac{V_{mis}(\hat{\alpha})}{V_{true}(\hat{\alpha})}=\frac{A'_{11}-A'_{12}-A'_{13}-\cdots-A'_{1,k+2}}{A_{11}-A_{12}-A_{13}-\cdots-A_{1,k+2}}\,. \qquad (4.2.27)$$

and

$$\frac{V_{mis}(\hat{\beta}_j)}{V_{true}(\hat{\beta}_j)}$$
$$=\frac{(A_{11}-A_{12}-\cdots-A_{1j}-A_{1,j+2}-\cdots-A_{1,k+2})(A'_{11}-A'_{12}-A'_{13}-\cdots-A'_{1,k+2})A'_{1j}}{(A'_{11}-A'_{12}-\cdots A'_{1j}-A'_{1,j+2}-\cdots-A'_{1,k+2})(A_{11}-A_{12}-A_{13}-\cdots-A_{1,k+2})A_{1j}},$$
(4.2.28)

for $j=1,2,\cdots,k+1$.

From (4.2.27) we have

$$
\begin{aligned}
&(A_{11} - A_{12} - A_{13} - \cdots - A_{1,k+2}) - (A'_{11} - A'_{12} - A'_{13} - \cdots - A'_{1,k+2}) \\
&= \sum_{i=1}^{n} e^{\hat{\alpha}+x_{i1}\hat{\beta}_1+x_{i2}\hat{\beta}_2+\cdots+x_{ik}\hat{\beta}_k+x_{i,k+1}\hat{\beta}_{k+1}}(t_i - t'_i)(1 - x_{i1} - x_{i2} - \cdots - x_{ik} - x_{i,k+1}) \\
&= \Big[\sum_{i=1}^{n} e^{\hat{\alpha}+x_{i1}\hat{\beta}_1+x_{i2}\hat{\beta}_2+\cdots+x_{ik}\hat{\beta}_k}(t_i - t'_i)(1 - x_{i1} - x_{i2} - \cdots - x_{ik})\Big]\Big|\{i : x_{i,k+1=0}\} \\
&= \sum_{i_{00\cdots0}} e^{\hat{\alpha}}(t_i - t'_i)\Big|\{i : x_{i,k+1=0}\} \quad \text{[by the assumption]} \\
&= \sum_{i_{00\cdots00}} e^{\hat{\alpha}}(t_i - t'_i) > 0,
\end{aligned}
$$

where $i_{00\cdots00} = \{i : x_{i1} = 0, x_{i2} = 0, \cdots, x_{ik} = 0, x_{i,k+1} = 0\}$.

Also from (4.2.28) it is easy to show

$$
\begin{aligned}
&(A'_{11}-A'_{12}-\cdots-A'_{1j}-A'_{1,j+2}-\cdots-A'_{1,k+2})(A_{11}-A_{12}-A_{13}-\cdots-A_{1,k+2})A_{1j} \\
&- (A_{11} - A_{12} - \cdots - A_{1j} - A_{1,j+2} - \cdots - A_{1,k+2})(A'_{11} - A'_{12} - A'_{13} - \cdots - A'_{1,k+2})A'_{1j} \\
&= \Big[(A'_{11}-A'_{12}-\cdots-A'_{1j}-A'_{1,j+2}-\cdots-A'_{1,k+1})(A_{11}-A_{12}-A_{13}-\cdots-A_{1,k+1})A_{1j} \\
&-(A_{11}-A_{12}-\cdots-A_{1j}-A_{1,j+2}-\cdots-A_{1,k+1})(A'_{11}-A'_{12}-A'_{13}-\cdots-A'_{1,k+1})A'_{1j}\Big]\Big|\{i : x_{i,k+1} = 0\} \\
&= \Big[\sum_{i_{00\cdots10\cdots0}} e^{\hat{\alpha}+\hat{\beta}_j}t_i \sum_{i_{00\cdots10\cdots0}} e^{\hat{\alpha}+\hat{\beta}_j}t'_i \sum_{i_{00\cdots0}} e^{\hat{\alpha}}(t_i - t'_i) \\
&+ \sum_{i_{00\cdots0}} e^{\hat{\alpha}}t_i \sum_{i_{00\cdots0}} e^{\hat{\alpha}}t'_i \sum_{i_{00\cdots10\cdots0}} e^{\hat{\alpha}+\hat{\beta}_j}(t_i - t'_i)\Big]\Big|\{i : x_{i,k+1} = 0\} \quad \text{[by the assumption]} \\
&= \sum_{i_{00\cdots10\cdots00}} e^{\hat{\alpha}+\hat{\beta}_j}t_i \sum_{i_{00\cdots10\cdots00}} e^{\hat{\alpha}+\hat{\beta}_j}t'_i \sum_{i_{00\cdots00}} e^{\hat{\alpha}}(t_i-t'_i)+\sum_{i_{00\cdots00}} e^{\hat{\alpha}}t_i \sum_{i_{00\cdots00}} e^{\hat{\alpha}}t'_i \sum_{i_{00\cdots10\cdots00}} e^{\hat{\alpha}+\hat{\beta}_j}(t_i-t'_i) > 0.
\end{aligned}
$$

From the Principle of Induction, we have proved that $V_{mis}(\hat{\alpha}) < V_{true}(\hat{\alpha})$ and $V_{mis}(\hat{\beta}_j) < V_{true}(\hat{\beta}_j)$ hold for the categorical covariates case.

Suppose p becomes large in such a way that $p \leq n$. Then we need to guarantee that the observed $(p + 1) \times (p + 1)$ information matrices (including $x_{i0} \equiv 1$) are invertible. This will be the case if they are of full column rank. But in the case of a categorical variable, the structure of the information matrix involved comprises a principal diagonal, a first row and a first column (see next section). Such a matrix may be shown to be of full column rank (and hence possess an inverse which can in this case be calculated in closed form) provided there is at least one observation per category. We note in passing that any continuous covariate can be represented in $p \leq n$ categories and hence for such a representation of a continuous covariate we have $V_{mis}(\hat{\alpha}) < V_{true}(\hat{\alpha})$ and $V_{mis}(\hat{\beta}_j) < V_{true}(\hat{\beta}_j), j = 1, 2, \cdots, p$.

4.3 Information Matrix

4.3.1 Derivation of Information Matrix

In Sections 3.7 and 3.8, only the observed information matrix $I^{[o]}(\alpha, \boldsymbol{\beta})$ was found. In this section, let us try to find the expected information matrix $\mathcal{I}(\alpha, \boldsymbol{\beta})$.

The primary use of $\mathcal{I}(\alpha, \boldsymbol{\beta})$ is for design purpose; the matrix $\mathcal{I}^{-1}(\alpha, \boldsymbol{\beta})$ at specified values of α and $\boldsymbol{\beta}$ can be used to estimate the precision of estimators based on a given sample size and censoring pattern.

For the sake of simplicity, let us denote the vector of parameters, α and $\boldsymbol{\beta}$, by $\boldsymbol{\theta}$, so that $\hat{\boldsymbol{\theta}}$ is a vector of maximum likelihood estimates. Let us also denote by $U(\boldsymbol{\theta})$ the vector of the first derivatives (called the efficient scores), we know that $U(\boldsymbol{\theta})$ is asymptotically distributed as $N\left[0, \mathcal{I}^{-1}(\boldsymbol{\theta})\right]$. The expected information matrix, $\mathcal{I}(\boldsymbol{\theta})$, is such that the (j, k)-th element is

$$\mathcal{I}(\boldsymbol{\theta}) = -E\left(\frac{\partial^2 \ell(\cdot)}{\partial\theta_j \partial\theta_k}\right).$$

For the observed information matrix $I^{[o]}(\boldsymbol{\theta})$ evaluated at $\hat{\boldsymbol{\theta}}$ and the expected information matrix $\mathcal{I}(\boldsymbol{\theta})$, we do know that

$$\lim_{n\to\infty} E\left[I^{[o]}(\hat{\boldsymbol{\theta}})\right] = \mathcal{I}(\boldsymbol{\theta})$$

and

$$\lim_{n\to\infty} V\left[I^{[o]}(\hat{\boldsymbol{\theta}})\right] = 0.$$

In other words, $I^{[o]}(\hat{\boldsymbol{\theta}})/n$ is a consistent estimator of $\mathcal{I}(\boldsymbol{\theta})/n$.

Usually, the Fisher information matrix $\mathcal{I}(\boldsymbol{\theta}) = E\left[I^{[o]}(\boldsymbol{\theta})\right]$ is not available unless the censoring process is fully specified, but, we know that it is not possible to observe fixed censoring time for all individuals when interval censoring is present in the data. For this reason, we can find the observed information matrix instead. Actually, we have found them. However, for the exponential regression model we still want to find a form of expected information matrix $\mathcal{I}(\boldsymbol{\theta})$.

Let us look back to the second partial derivatives obtained in the Section 3.8 in Chapter 3. For simplicity, we would consider (3.8.27), (3.8.28) and (3.8.29), which were got by using the first order term approximation of R_i. Taking the expectation on these three second derivatives, we have the expected information matrix in form:

$$\mathcal{I}(\alpha, \boldsymbol{\beta}) = \begin{bmatrix} \sum_{i=1}^{n} \mathrm{e}^{\alpha+\mathbf{x}_i'\boldsymbol{\beta}} E(t_i') & \sum_{i=1}^{n} \mathbf{x}_i' \mathrm{e}^{\alpha+\mathbf{x}_i'\boldsymbol{\beta}} E(t_i') \\ \sum_{i=1}^{n} \mathbf{x}_i \mathrm{e}^{\alpha+\mathbf{x}_i'\boldsymbol{\beta}} E(t_i') & \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' \mathrm{e}^{\alpha+\mathbf{x}_i'\boldsymbol{\beta}} E(t_i') \end{bmatrix},$$

where $E(t_i') = E\left[\delta_i t_{i,k-1}^* + (1-\delta_i) t_i^c\right]$, $t_{i,k-1}^*$ are fixed, δ_i and t_i^c are linked to the specified censoring process. Clearly, it is not easy to obtain this expectation. However, since t_i' implies any time, so we may let it be some time t_i, then

$$E(t_i') = \mathrm{mean}(t_i) = \frac{1}{\lambda(t_i)} = \mathrm{e}^{-(\alpha+\mathbf{x}_i'\boldsymbol{\beta})},$$

therefore

$$\mathcal{I}(\alpha, \boldsymbol{\beta}) = \begin{bmatrix} n & \sum_{i=1}^n \mathbf{x}_i' \\ \sum_{i=1}^n \mathbf{x}_i & \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' \end{bmatrix}.$$

Especially, if $\mathbf{x}_i$ is p categorical covariates, we have

$$\mathcal{I}(\alpha, \boldsymbol{\beta}) = \begin{bmatrix} n & n_1 & n_2 & \cdots & n_p \\ n_1 & n_1 & 0 & \cdots & 0 \\ n_2 & 0 & n_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ n_p & 0 & 0 & \cdots & n_p \end{bmatrix},$$

where $n_j = \sum_{i=1}^n x_{ij}$ for $j = 1, 2, \cdots, p$. The inverse of $\mathcal{I}(\alpha, \boldsymbol{\beta})$ above, after some direct algebras, is

$$\Big[\mathcal{I}(\alpha, \boldsymbol{\beta})\Big]^{-1} = \frac{1}{(\prod_{j=1}^p n_j)(n - \sum_{j=1}^p n_j)} \times$$

$$\begin{bmatrix} \prod_{j=1}^p n_j & -\prod_{j=1}^p n_j & -\prod_{j=1}^p n_j & \cdots & -\prod_{j=1}^p n_j \\ -\prod_{j=1}^p n_j & (\prod_{j\neq 1} n_j)(n - \sum_{j\neq 1} n_j) & \prod_{j=1}^p n_j & \cdots & \prod_{j=1}^p n_j \\ -\prod_{j=1}^p n_j & \prod_{j=1}^p n_j & (\prod_{j\neq 2} n_j)(n - \sum_{j\neq 2} n_j) & \cdots & \prod_{j=1}^p n_j \\ \vdots & \vdots & \vdots & & \vdots \\ -\prod_{j=1}^p n_j & \prod_{j=1}^p n_j & \prod_{j=1}^p n_j & \cdots & (\prod_{j\neq p} n_j)(n - \sum_{j\neq p} n_j) \end{bmatrix} \tag{4.3.29}$$

When $p = 1$, the (4.3.29) becomes

$$\mathcal{I}^{-1}(\alpha, \beta) = \frac{1}{nn_1 - n_1^2} \begin{bmatrix} n_1 & -n_1 \\ -n_1 & n \end{bmatrix}.$$

Thus

$$V(\alpha) = \frac{n_1}{nn_1 - n_1^2} = \frac{1}{n(1 - \frac{n_1}{n})}, \tag{4.3.30}$$

$$V(\beta) = \frac{n}{nn_1 - n_1^2} = \frac{1}{n_1(1 - \frac{n_1}{n})}\ . \tag{4.3.31}$$

The above analysis is an approximation and other lines of attack are possible in which $E(t_i') = E\big[\delta_i t_{i,k-1}^* + (1-\delta_i)t_i^c\big]$ where this time $E\big[\delta_i t_{i,k-1}^*\big]$ is treated as a constant (i.e., as part of the fixed schedule) and $E\big[(1-\delta_i)t_i^c\big]$ is replaced by the future expectation of the censored exponential random variable, i.e., by $(1-\delta_i)[t_i^c + 1/\lambda(t_i)]$ whence $\mathcal{I} = \mathcal{I}_u + \mathcal{I}_c$.

4.3.2 Hypothesis Testing

Lets now discuss possible hypothesis testing procedures. Usually, we would be interested in testing

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

Popular hypothesis tests are the likelihood ratio test, score test and Wald test, and we now describe each one briefly.

In general, a likelihood ratio test is test given by the critical region C of the form

$$C = \left\{ X : \lambda = \frac{L(H_0)}{L(H_1)} < k \right\},$$

for some constant k (the value of k is determined by fixing the size α of the test, so that $P(X \in C|\ H_0 = \alpha)$). The likelihood ratio test statistic in our case is

$$-2\log\left(\frac{L(\boldsymbol{\theta}_0)}{\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta})}\right)$$

and we know that this is asymptotically distributed as χ_p^2 under H_0.

Both Wald test and score test assume that $\hat{\boldsymbol{\theta}} \sim N\big[\boldsymbol{\theta}, I^{-1}(\boldsymbol{\theta})\big]$. The Wald test statistic is

$$W = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\big[I^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\big](\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$$

if $H_0 : \boldsymbol{\theta} = \mathbf{0}$ then $W = \hat{\boldsymbol{\theta}}'\big[I^{-1}(\hat{\boldsymbol{\theta}})\big]\hat{\boldsymbol{\theta}}$ and under H_0, W is asymptotically distributed as χ^2_p.

The Score test statistic is

$$W = U'(\hat{\boldsymbol{\theta}})\big[I^{-1}(\hat{\boldsymbol{\theta}})\big]U(\hat{\boldsymbol{\theta}})$$

which is also asymptotically distributed as χ^2_p under H_0.

In all three tests, the calculated value of the test statistic would be compared to the tabulated value, and rejection or non-rejection of H_0 would follow accordingly.

We would like be very interested in comparing the pseudo and the true likelihood with regard to the significance of parameters. This analysis has been carried out in the simulation study and the real data study.

Chapter 5

Simulation Study

Simulation is a technique for performing sampling experiments on the model of the system. The experiments are calculated using a computer model rather than on the real system as the latter would be inconvenient, expensive and time consuming. The simulated experiments should be based on statistical theory (Hillier and Lieberman (1995)).

In this thesis, a simulation study was conducted to evaluate the finite sample performance of the pseudo and true models by estimating the bias and precision of the MLEs in the exponential distribution with covariate. The simulation will be data directed based on an age-related macular disease (ARMD) study by Hart et al. (2002). The goal is to quantify the degree to which inference about the parameters differs in the models, especially to check the results of the analytical work obtained in Section 4.2, and the relationship between the observed information and the expected information in Section 4.3 in the interval censored setting.

5.1 Simulation Settings

5.1.1 Generation of the Exponential Regression Model

Simulation involves generating vector $u = (u_1, u_2, \cdots, u_n)$ using random number generator with $U_i \sim \text{Uniform}(0,1)$ and then using the inverse function to find $t = (t_1, t_2, \cdots, t_n)$ as follows

$$F(t; \alpha, \beta, x) = 1 - S(t; \alpha, \beta, x),$$

where $F(t; \alpha, \beta, x) = u \sim \text{Uniform}(0,1)$ and $S(t; \alpha, \beta, x)$ is the survival function of exponential regression model with parameters α and β and covariate x. Then

$$u_i = 1 - \exp(-\mathrm{e}^{\alpha + x_i\beta} t_i),$$

from which

$$t_i = \frac{-\log_{\mathrm{e}}(1 - u_i)}{\mathrm{e}^{\alpha + x_i\beta}}\ .$$

Thus, we obtain a random sample of observations from the density of the exponential regression with pre-specified values of parameters α and β and $x_i, i = 1, 2, \cdots, n$, is set to take values 0 and 1, say. We then use both the pseudo and the true likelihood to analyze the data. By varying the sample size, percentage of observations right censored, the range of parameters, the scheduled inspection times and the points (mid or end) of the interval, we can study how inference on parameters is affected by the use of a pseudo-likelihood.

5.1.2 Parametric Settings

To set the simulation, we should handle some key points as follows:

1. **Dealing with the inspection times**

The object (3.1.1) shown in Chapter 3 is called the schedule, which fixes the inspection times of patients. It is all idealistic object. Patients are humans and do not observe the schedule, in reality, every patient has his own pattern of attendance in the study. Many may come a few days/weeks earlier or later, e.g. ARMD study. Finkelstein (1986) argued that the same fixed follow-up schedule was required for each individual and this approach was echoed by Collett (2003). However, this approach does not respect everyday experience in clinical trials where, typically, individuals do not observe the fixed follow-up schedule. Accordingly, this restrictive assumption was relaxed. We have allowed for this in our simulation study, by adding a normally distributed component to each inspection time. For the i-th individual, the actual visit time will be (schedule time + e_i) rather than the schedule. Where e_i is normally distributed about 0 with standard deviation σ determined as follows. A maximum error of $\pm$ month was allowed. Thus, using the normal distribution, $6\sigma = 2$ months and therefore $\sigma = 1/3$ of a month. So, $e_i \sim N(0, 1/3)$ where $i = 1, 2, \cdots, n$. In this way, for the schedule (3, 6, 12, 24) in ARMD, the simulated visit time is $(3, 6, 12, 24) + e_i$. The generated failure times t_i were classified into regular and irregular intervals as defined above.

As we have seen in Section 3.5, the relative performance of the estimators from the true likelihood depends on the widths of the intervals between inspection times. Therefore, we vary the frequency and regularity of inspection times within the 24 months of follow up period.

2. **Control of censoring percentage and selection of parameters**

We let percentage of right censored times varied since it too has impact on inference. We look at 0.05, 0.10 and 0.30. Let us now explain briefly how, for example, 10 percent right censoring was generated.

Consider the two sample case, i.e. $x_i = 0$ or 1, we have

$$S(t_i; \alpha, \beta, x_i) = \exp(-\mathrm{e}^{\alpha + x_i \beta} t_i).$$

For a fixed time t_c, if $t_i \le t_c$, we let the corresponding t_i be uncensored (meaning an event in pseudo-likelihood, and interval censored in the true likelihood) so that the indicator $\delta_i = 1$. If $t_i > t_c$, the corresponding t_i is said to be right censored (i.e. $\delta_i = 0$). Then the probability of right censored p is

$$p = S(t_c; \alpha, \beta, x_i) = \exp(-\mathrm{e}^{\alpha + x_i \beta} t_c).$$

Let $p = 0.10$, we have

$$0.10 = \exp(-\mathrm{e}^{\alpha + 0 \times \beta} t_c)$$
$$0.10 = \exp(-\mathrm{e}^{\alpha + 1 \times \beta} t_c),$$

then

$$0.20 = \exp(-\mathrm{e}^{\alpha + 0 \times \beta} t_c) + \exp(-\mathrm{e}^{\alpha + 1 \times \beta} t_c). \tag{5.1.1}$$

Thus, for the selected value of β we can get the corresponding value of α by solving equation (5.1.1). This procedure can never guarantee exactly 10 percent of right censoring, but in all cases it will be very close. For our data directed simulation, $t_c = 24$ is used, where we assumed a Type I censoring process.

3. Technical settings for the true likelihood

Let $\hat{\alpha}^{(0)}$ be $\log\left[\frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} \mathrm{e}^{x_i \beta} t_i}\right]$, the approximate (and an exact in the case of a pseudo-likelihood) maximum likelihood estimate of α; and let $\hat{\alpha}$ be the true maximum likelihood estimate of α. This notation is somewhat inconsistent with that used so far, but will serve the purpose of providing clearer understanding.

Lets expand $\frac{\partial \ell_{true}(\hat{\alpha},\beta)}{\partial \alpha}$ about $\hat{\alpha}^{(0)}$,

$$\frac{\partial \ell_{true}(\hat{\alpha},\beta)}{\partial \alpha} = \frac{\partial \ell_{true}(\hat{\alpha}^{(0)},\beta)}{\partial \alpha} + \frac{\partial \ell_{true}}{\partial \alpha}\left[\frac{\partial \ell_{true}(\hat{\alpha}^{(0)},\beta)}{\partial \alpha}\right]\left[\hat{\alpha} - \hat{\alpha}^{(0)}\right] + \frac{1}{2}\frac{\partial^2 \ell_{true}}{\partial \alpha^2}\left[\frac{\partial \ell_{true}(\hat{\alpha}^{(0)},\beta)}{\partial \alpha}\right]\left[\hat{\alpha} - \hat{\alpha}^{(0)}\right]^2 + \cdots \quad (5.1.2)$$

we know that α satisfies $\frac{\partial \ell_{true}(\hat{\alpha},\beta)}{\partial \alpha} = 0$, so

$$\frac{\partial \ell_{true}(\hat{\alpha}^{(0)},\beta)}{\partial \alpha} + \sum_{i=1}^{n} \frac{1}{n!}\frac{\partial^n \ell_{true}}{\partial \alpha^n}\left[\frac{\partial \ell_{true}(\hat{\alpha}^{(0)},\beta)}{\partial \alpha}\right]\left[\hat{\alpha} - \hat{\alpha}^{(0)}\right]^n = 0,$$

from this, we would need to express $\hat{\alpha}$ as

$$\hat{\alpha} = \hat{\alpha}^{(0)} - B,$$

where B is bias. But it is not possible to see if (5.1.2) is invertible, so we cannot obtain the bias in its closed form. Therefore, we make the following approximation, on which the Newton-Raphson iterative method is based

$$\frac{\partial \ell_{true}(\hat{\alpha},\beta)}{\partial \alpha} \simeq \frac{\partial \ell_{true}(\hat{\alpha}^{(0)},\beta)}{\partial \alpha} + \frac{\partial \ell_{true}}{\partial \alpha}\left[\frac{\partial \ell_{true}(\hat{\alpha}^{(0)},\beta)}{\partial \alpha}\right]\left[\hat{\alpha} - \hat{\alpha}^{(0)}\right],$$

from which

$$\hat{\alpha} = \hat{\alpha}^{(0)} - \frac{\frac{\partial \ell_{true}(\hat{\alpha}^{(0)},\beta)}{\partial \alpha}}{\frac{\partial^2 \ell_{true}(\hat{\alpha}^{(0)},\beta)}{\partial \alpha^2}} \quad (5.1.3)$$

This is the iterative process of the Newton-Raphson (N-R) method. The RHS of (5.1.3) will yield the second approximation, $\alpha^{(1)}$, to α. The iterative process, with each step using an improved estimate, is continued until we get desired accuracy, i.e. until our approximations converge to $\hat{\alpha}$, which could be checked by assessing the relative change in the log-likelihood. A useful "bonus" of the N-R method is that, upon convergence, the approximate variance-covariance matrix of parameter estimates can be obtained by inverting the corresponding observed information matrix.

The last term of (5.1.3) is the approximate bias for α which can be found by evaluating the derivatives (already found) at $\hat{\alpha}^{(0)}$. Similarly, we can obtained the approximate expression between $\hat{\beta}$ and $\hat{\beta}^{(0)}$.

However, caution is required when employing N-R method to the true likelihood. Several requirements are needed to ensure that we get true convergence. Firstly, the starting points $\left(\hat{\alpha}^{(0)}, \hat{\beta}^{(0)}\right)$ should be reasonably close to $(\hat{\alpha}, \hat{\beta})$. Since $\left(\hat{\alpha}^{(0)}, \hat{\beta}^{(0)}\right)$ are the same as the exact estimators from pseudo-likelihood $(\hat{\alpha}, \hat{\beta})_{mis}$, then we used $(\hat{\alpha}, \hat{\beta})_{mis}$ as the starting points to the $(\hat{\alpha}, \hat{\beta})$ from true likelihood. And secondly, the log-likelihood should be approximately quadratic in vicinity of $(\hat{\alpha}, \hat{\beta})$. Otherwise, we may observe convergence to a local maximum, or not observe convergence at all.

In addition, to avoid infinity appearing in the log true likelihood due to $1 - S(t^*_{ik})/S(t^*_{i,k-1}) \simeq 0$, we set $S(t^*_{ik})/S(t^*_{i,k-1}) = 0.999$ a constant when $S(t^*_{ik})/S(t^*_{i,k-1}) > 0.999$ in the simulation.

Based on the things involved in the simulation discussed above, the following simulation settings were chosen to be varied simultaneously

(a) Sample size: $n = 100, 200, 500$;

(b) Approximate percentage of right censored times: $p = 0.05, 0.10, 0.30$;

(c) Scheduled inspection times (in months): (3,6,12,24) and (3,6,9,12,15,18,21,24);

(d) $\lambda_0 : 0.1$ or $\alpha = \log(\lambda_0) : -2.3026$;

(e) $\beta : 0.668, -0.0807, -1.27$.

5.1.3 Computational Methods

As mentioned previously, end points or mid points of the observed intervals are treated as if they were exact failure times and are analyzed using the pseudo-likelihood. The results are compared with those obtained using the true likelihood which utilizes the interval information directly. Interest was focused on the treatment effect β rather than the scale parameter $\lambda_0 = \exp(\alpha)$.

Therefore, the use of these two likelihoods enables one investigate the effect of mis-specification for any survival model which has a closed form survivor function. In this situation, the exponential regression distribution was investigated. The observed inspection times rather than the scheduled times were used in both likelihoods.

As a simulation size of $N = 1000$ iterations was used, the estimates of the parameter α and β were expected to be Normally distributed, according to the classical asymptotic theory. Thus, one may compare the estimate of the parameter and its variances between pseudo and the true likelihood models.

The variances were compared by computing the ratio, of the variance obtained from the pseudo-likelihood and the true likelihood. When the ratio is less than 1, the variance obtained from the pseudo likelihood is less than that obtained from the true likelihood, i.e., it is artificially precise.

The Mean Square Error (MSE) for α, β was calculated as follows

$$\text{MSE}(\alpha, \beta) = \text{Var}(\alpha, \beta) + \left(\text{Bias}(\alpha, \beta)\right)^2,$$

where the bias is defined as

$$\text{Bias}(\alpha, \beta) = \left(\bar{\hat{\alpha}}, \bar{\hat{\beta}}\right) - (\alpha, \beta),$$

where $\bar{\hat{\alpha}}$ and $\bar{\hat{\beta}}$ are expectations of MLEs. If, $\text{MSE}(\alpha, \beta) = \text{Var}(\alpha, \beta)$, then the bias is zero, otherwise the bias was calculated as

$$\text{Bias}(\alpha, \beta) = \sqrt{\left(\text{Bias}(\alpha, \beta)\right)^2} = \sqrt{\text{MSE}(\alpha, \beta) - \text{Var}(\alpha, \beta)}.$$

5.2 Results of Simulation

Programmes for simulations were created in R and one such program is presented in Appendix A where the procedure for calculating different components is explained in detail. The results of simulation with the end points used in the pseudo-ikelihood are listed in Table 5.1, Table 5.2, Table 5.3 and Table 5.4.

5.3 Findings

We already know that, in general, the performance of estimators is better with larger sample size and smaller percentage of right censored times. This is true for the estimate on β in our investigation, but for α, its performance was not affected by the censoring rate due to our regulation of censoring rate control.

It seemed reasonable to assess the performance of estimators within each schedule of inspection times in order to get more insight into the effects of the frequency and regularity of inspection times. And within the two schedules, it was decided to investigate not the absolute performance of the MLEs from the two likelihoods, though this is certainly of great importance too, but the difference between the two.

For the sake of outlining the adopted approach as clearly as possible, let us consider the assessment of MLEs of α from the two likelihoods for the two schedules. To this

Table 5.1: *The maximum likelihood estimates when* $p = 0.05, \lambda_0 = 0.1$ *or* $\alpha = log(\lambda_0) = -2.3026, \beta = 0.668$.

Values	Scheduled inspection times (in months)					
	(3, 6, 12, 24)			(3, 6, 9, 12, 15, 18, 21, 24)		
n	100	200	500	100	200	500
$\bar{\hat{\alpha}}_{mis}$	-2.5876	-2.5895	-2.5938	-2.4486	-2.4515	-2.4500
$\bar{\hat{\alpha}}_{true}$	-2.2945	-2.2980	-2.3040	-2.2969	-2.3012	-2.3000
$\bar{\hat{\beta}}_{mis}$	0.5385	0.5399	0.5421	0.5385	0.5399	0.5319
$\bar{\hat{\beta}}_{true}$	0.6677	0.6696	0.6718	0.6741	0.6773	0.6661
$V_{mis}(\hat{\alpha})$	0.0220	0.0110	0.0044	0.0220	0.0110	0.0044
$V_{true}(\hat{\alpha})$	0.0229	0.0114	0.0046	0.0222	0.0111	0.0044
$V_{mis}(\hat{\beta})$	0.0422	0.0211	0.0084	0.0422	0.0211	0.0084
$V_{true}(\hat{\beta})$	0.0448	0.0224	0.0089	0.0430	0.0215	0.0086
$\text{MSE}_{mis}(\hat{\alpha})$	0.1032	0.0933	0.0892	0.0433	0.0332	0.0261
$\text{MSE}_{true}(\hat{\alpha})$	0.0230	0.0115	0.0046	0.0222	0.0111	0.0044
$\text{MSE}_{mis}(\hat{\beta})$	0.0589	0.0374	0.0242	0.0594	0.0374	0.0269
$\text{MSE}_{true}(\hat{\beta})$	0.0448	0.0224	0.0089	0.0431	0.0216	0.0086

aim, we wish to test the hypotheses $H_0 : E(\hat{\alpha}_{true}) = E(\hat{\alpha}_{mis})$ vs. $H_a : E(\hat{\alpha}_{true}) \neq E(\hat{\alpha}_{mis})$. We first define the difference as $\bar{\hat{\alpha}}_{true} - \bar{\hat{\alpha}}_{mis}$, where $\bar{\hat{\alpha}}_{true}$ is the mean of the 1000×1 vector of true likelihood values of α for each simulation of the two schedules in Table 5.1, Table 5.2 and Table 5.3. Similarly for $\bar{\hat{\alpha}}_{mis}$. We then perform a paired t-test on this difference to see whether there is sufficient evidence for or against the null hypothesis of zero difference.

Since all the p-values with respect to the 18 paired t-tests (at 5 percent significance level) on the difference between the MLEs of α from the two likelihoods are less than 2.2e^{-16}, we can see that the null hypothesis: $H_0 : E(\hat{\alpha}_{true}) = E(\hat{\alpha}_{mis})$ was rejected.

Table 5.2: *The maximum likelihood estimates when* $p = 0.10, \lambda_0 = 0.1$ *or* $\alpha = log(\lambda_0) = -2.3026, \beta = -0.0807$.

Values	Scheduled inspection times (in months)					
	(3, 6, 12, 24)			(3, 6, 9, 12, 15, 18, 21, 24)		
n	100	200	500	100	200	500
$\bar{\hat{\alpha}}_{mis}$	-2.5876	-2.5895	-2.5938	-2.4486	-2.4515	-2.4500
$\bar{\hat{\alpha}}_{true}$	-2.2945	-2.2980	-2.3040	-2.2969	-2.3012	-2.3000
$\bar{\hat{\beta}}_{mis}$	-0.0702	-0.0652	-0.0633	-0.0693	-0.0651	-0.0726
$\bar{\hat{\beta}}_{true}$	-0.0867	-0.0809	-0.0776	-0.0806	-0.0758	-0.0844
$V_{mis}(\hat{\alpha})$	0.0220	0.0110	0.0044	0.0220	0.0110	0.0044
$V_{true}(\hat{\alpha})$	0.0229	0.0114	0.0046	0.0222	0.0111	0.0044
$V_{mis}(\hat{\beta})$	0.0446	0.0223	0.0089	0.0445	0.0222	0.0089
$V_{true}(\hat{\beta})$	0.0463	0.0231	0.0092	0.0449	0.0224	0.0090
$\text{MSE}_{mis}(\hat{\alpha})$	0.1032	0.0933	0.0892	0.0433	0.0332	0.0261
$\text{MSE}_{true}(\hat{\alpha})$	0.0230	0.0115	0.0046	0.0222	0.0111	0.0044
$\text{MSE}_{mis}(\hat{\beta})$	0.0447	0.0225	0.0092	0.0447	0.0225	0.0090
$\text{MSE}_{true}(\hat{\beta})$	0.0463	0.0231	0.0092	0.0449	0.0224	0.0090

We concluded that there is sufficient evidence to suggest that there is a difference between $E(\hat{\alpha}_{true})$ and $E(\hat{\alpha}_{mis})$.

The procedure is the same for the two schedules, as well as for analyzing $H_0 : E(\hat{\beta}_{true}) = E(\hat{\beta}_{mis})$. We found that there is also sufficient evidence to suggest that there is a difference between $E(\hat{\beta}_{true})$ and $E(\hat{\beta}_{mis})$ since all the p-values with respect to the 18 paired t-tests (at 5 percent significance level) on the difference between the MLEs of β from the two likelihoods are also less than 2.2e^{-16}.

Another important comment that needs to be made is that of lack of consistency shown by $\hat{\alpha}_{mis}$ as estimator of α. It is clear from both schedules, that $\hat{\alpha}_{mis}$ does not

Table 5.3: *The maximum likelihood estimates when* $p = 0.30, \lambda_0 = 0.1$ *or* $\alpha = log(\lambda_0) = -2.3026, \beta = -1.27$.

Values	Scheduled inspection times (in months)					
	(3, 6, 12, 24)			(3, 6, 9, 12, 15, 18, 21, 24)		
n	100	200	500	100	200	500
$\bar{\hat{\alpha}}_{mis}$	-2.5876	-2.5895	-2.5938	-2.4486	-2.4515	-2.4500
$\bar{\hat{\alpha}}_{true}$	-2.2945	-2.2980	-2.3040	-2.2969	-2.3012	-2.2300
$\bar{\hat{\beta}}_{mis}$	-1.1004	-1.0885	-1.0839	-1.1728	-1.1644	-1.1684
$\bar{\hat{\beta}}_{true}$	-1.2868	-1.2736	-1.2677	-1.2810	-1.2715	-1.2757
$V_{mis}(\hat{\alpha})$	0.0220	0.0110	0.0044	0.0220	0.0110	0.0044
$V_{true}(\hat{\alpha})$	0.0229	0.0114	0.0046	0.0222	0.0111	0.0044
$V_{mis}(\hat{\beta})$	0.0637	0.0315	0.0126	0.0636	0.0316	0.0126
$V_{true}(\hat{\beta})$	0.0647	0.0321	0.0128	0.0636	0.0317	0.0126
$\text{MSE}_{mis}(\hat{\alpha})$	0.1032	0.0933	0.0892	0.0433	0.0332	0.0261
$\text{MSE}_{true}(\hat{\alpha})$	0.0230	0.0115	0.0046	0.0222	0.0111	0.0044
$\text{MSE}_{mis}(\hat{\beta})$	0.0921	0.0641	0.0468	0.0729	0.0425	0.0227
$\text{MSE}_{true}(\hat{\beta})$	0.0651	0.0321	0.0128	0.0640	0.0317	0.0127

approch α as n increases.

In Section 4.2, we have shown that the MLEs from pseudo-likelihood have smaller variance than MLEs from the true likelihood. Here we also have $\frac{V_{mis}(\hat{\alpha})}{V_{true}(\hat{\alpha})} < 1$ and $\frac{V_{mis}(\hat{\beta})}{V_{true}(\hat{\beta})} < 1$. Further, one sample t-tests (5 percent significance level) on $\log\left(\frac{V_{mis}(\hat{\alpha})}{V_{true}(\hat{\alpha})}\right)$ and $\log\left(\frac{V_{mis}(\hat{\beta})}{V_{true}(\hat{\beta})}\right)$ have been carried out for each schedule to check whether or not the differences between them are significant. The results are accordant to what was expected; there is significant evidence that the variances of MLEs from pseudo-likelihood are not equal to those from the true likelihood for both schedules, where the Null: $\log\left(\frac{V_{mis}(\hat{\alpha})}{V_{true}(\hat{\alpha})}\right) = 0$ or $\log\left(\frac{V_{mis}(\hat{\beta})}{V_{true}(\hat{\beta})}\right) = 0$ was rejected because all the p-values of these

Table 5.4: *Variance comparisons from the true likelihood and from Fisher information with large sample size.*

Values	Scheduled inspection times (in months)					
	(3, 6, 12, 24)			(3, 6, 9, 12, 15, 18, 21, 24)		
n	100	200	500	100	200	500
$V(\alpha)$	0.0200	0.0100	0.0040	0.0200	0.0100	0.0040
$(V_{true}(\hat{\alpha}) - V(\alpha))_{0.05}$	0.0029	0.0014	0.0006	0.0022	0.0011	0.0004
$(V_{true}(\hat{\alpha}) - V(\alpha))_{0.10}$	0.0029	0.0014	0.0006	0.0022	0.0011	0.0004
$(V_{true}(\hat{\alpha}) - V(\alpha))_{0.30}$	0.0029	0.0014	0.0006	0.0022	0.0011	0.0004
$V(\beta)$	0.0400	0.0200	0.0080	0.0400	0.0200	0.0080
$(V_{true}(\hat{\beta}) - V(\beta))_{0.05}$	0.0048	0.0024	0.0009	0.0030	0.0015	0.0006
$(V_{true}(\hat{\beta}) - V(\beta))_{0.10}$	0.0063	0.0031	0.0012	0.0049	0.0024	0.0010
$(V_{true}(\hat{\beta}) - V(\beta))_{0.30}$	0.0248	0.0121	0.0048	0.0238	0.0117	0.0046

tests are less than $2.2e^{-16}$.

Look at Table 5.1 to Table 5.3 again, we found that even though both ratios $\frac{V_{mis}(\hat{\alpha})}{V_{true}(\hat{\alpha})}$ and $\frac{V_{mis}(\hat{\beta})}{V_{true}(\hat{\beta})}$ are less than 1, but $\text{MSE}_{true}(\hat{\alpha}) < \text{MSE}_{mis}(\hat{\alpha})$ and $\text{MSE}_{true}(\hat{\beta}) < \text{MSE}_{mis}(\hat{\beta})$, and from (5.1.3) we know that $\text{Bias}(\alpha, \beta)$ from the true likelihood were less than that from the pseudo-likelihood which means that the MLEs from the true likelihood are more accurate. Throughout the simulation, Tables 5.1-5.3, the bias in the parameter (α, β) were relatively invariant to an increasing censoring rate in the true likelihood.

Figure 5.1 showed the distribution of the estimated parameters which are nearly normal ones. Look at Table 5.4, the estimated variances from the true likelihood are close to their theoretical values (obtained from Fisher information matrix by using the forms (4.3.30) and (4.3.31)). This confirmed that the expected information matrix

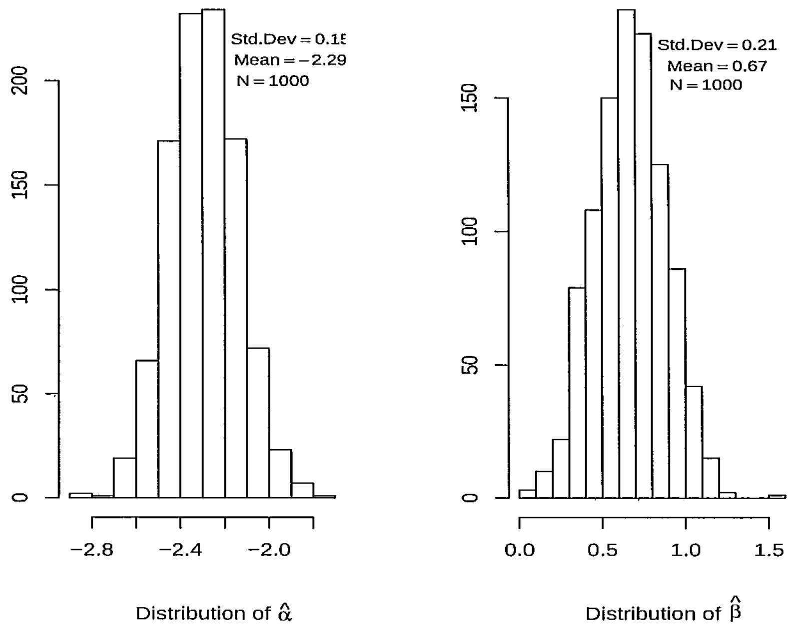

Figure 5.1: *The histograms of estimated parameters.*

derived in Section 4.3 is reasonable.

In conclusion, the results show that the conjecture, that the likelihood leads to over-precise estimates is true for the the exponential regression model and for both schedules, regular or irregular. The simulation results support the theoretical work that has been presented in Chapter 4.

There are strong evidence to show the significant difference between the MLEs and the variances from both likelihoods and for both schedules, regular or irregular. The tests relied on the usual standard errors available from a numerical estimate of the inverse of the observed information matrix.

From Table 5.1to Table 5.3 we noted that estimate under a frequent and regular schedule is better than a rare and irregular schedule.

It is, however, true that the use of the proposed form of the true likelihood, led, when the model performed satisfactorily, to results which were better in terms of precision and bias. Overall, the true likelihood is to be preferred.

Also we used mid points in the pseudo-likelihood to do the simulation, and found that all the conclusions are the same as those obtained by using the end points in the pseudo-likelihood above (The corresponding results are not shown in the thesis), but, estimators of pseudo-likelihood with the mid points used are more accurate than the estimators with the end points used.

Chapter 6

Application

In this chapter, we demonstrate the proposed estimation procedures by applying it to a set of interval-censored data obtained from an age-related macular disease (ARMD) study (Hart et al. (2002)). The original objective of the study was to determine whether teletherapy with 6-mV photons (Population 1, 101 patients) can reduce visual loss in patients with subfoveal choroidal nerovascularization in ARMD. Our R codes for computation are in Appendix B.

6.1 Description of Data

Two hundred and three patients were randomly assigned to radiotherapy or observation. Treatment was undertaken at designed radiotherapy centers and patients assigned to the treatment group received a total dosage of 12-Gy of 6-mV photons in 6 fractions. Follow-up was scheduled at 3, 6, 12 and 24 months. After excluding protocol violators, the data from 199 patients were analyzed.

The primary outcome measure was the loss of distance visual acuity in the study eye at 12 and 24 months. Other clinical outcomes of importance were the time to loss of 3 or more or 6 or more lines of distance visual acuity from baseline. Here we keep the same notations given in Hart et al. (2002), see Table 6.1.

Table 6.1: *Description of variables.*

Variables	Interpretation
tlost3	end of the interval in which 3 lines were lost
slost3	beginning of the interval in which 3 lines were lost
event3	event indicator (1=3 lines lost, 0= 3 lines not lost)
tlost6	end of the interval in which 6 lines were lost
slost6	beginning of the interval in which 6 lines were lost
event6	event indicator (1=6 lines lost, 0=6 lines not lost)
rx	treatment indicator (1=teletherapy, 0=control)

6.2 Results of Analysis

Step 1. Graphical checking for the distribution

Graphical method is an intuitive way to check whether the real data follow some parametric distribution. In survival analysis, the graphical checking is to obtain the Kaplan-Meier estimate $\hat{S}(t)$ and plot it against t or a suitable function of t, in order to see if it takes nearly the same shape as assumed model survivor function $S(t)$. Some diagnostic graphs can be derived for some distributions by examining the specific form of $S(t)$ in each case. Table 6.2 lists some of these. In all cases, $\hat{S}(t)$ refers to the model based estimator. We denote $\Phi(z)$ as the standard normal distribution function.

For example, for the exponential distribution, we know that

$$S(t) = \mathrm{e}^{-\lambda t} \qquad \text{and} \quad \ln S(t) = -\lambda t, \ t > 0, \tag{6.2.1}$$

this means that a scatter plot of the values of $\ln S(t)$ against $x = t$ should give a straight line if the assumption of an exponential distribution is correct.

Table 6.2: *Useful functions for graphic diagnosis for some distributions.*

Distributions	Functions		
Exponential	$\ln \hat{S}(t)$	with	t
Weibull	$\ln\{-\ln \hat{S}(t)\}$	with	$\ln t$
Gumbel	$\ln\{-\ln \hat{S}(t)\}$	with	t
Log-normal	$\Phi^{-1}(1-\hat{S}(t))$	with	$\ln t$
Gamma	$\Phi^{-1}(1-\hat{S}(t))$	with	$\sqrt{t}$
Normal	$\Phi^{-1}(1-\hat{S}(t))$	with	t
Logistic	$\ln\left(\frac{1-\hat{S}(t)}{\hat{S}(t)}\right)$	with	t
Log-logistic	$\ln\left(\frac{1-\hat{S}(t)}{\hat{S}(t)}\right)$	with	$\ln t$
Pareto	$\ln \hat{S}(t)$	with	$\ln t$

For the ARMD data, taking

$$\text{timd3} = \begin{cases} \frac{\text{tlost3}+\text{slost3}}{2} & \text{if event3} = 1 \\ \text{tlost3} & \text{if event3=0,} \end{cases}$$

where, we see that the mid points were used in the pseudo-likelihood since we know from the simulation that estimates at them are better. Making the Kaplan-Meier estimate on the pair (timd3, event3), we obtained the estimated survivor function st3 (Figure 6.1). To find a possible distribution for st3, some graphical checking was done. See Figure 6.2.

From Figure 6.2, we found that an exponential distribution for the data is a reasonable choice. A similar procedure was performed on the pair (timd6, event6), and the exponential distribution was selected again.

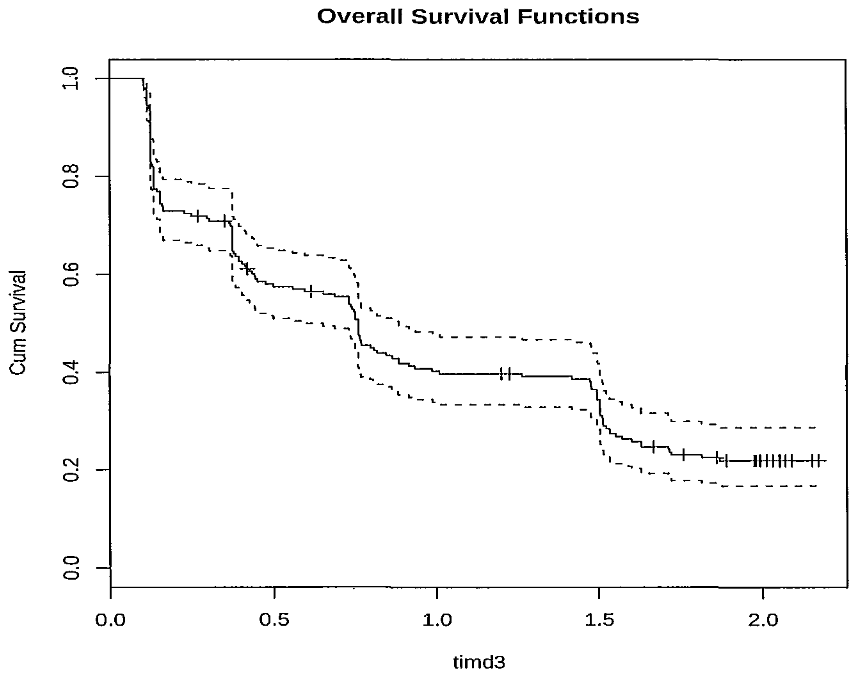

Figure 6.1: *The Kaplan-Meier Estimate Curve.*

Step 2. Parametric estimate

According to the results obtained in Chapter 3, we did the estimate by using the pseudo-likelihood to get the starting points for the estimation using the true likelihood, then we did the exact estimate. The results were listed in Table 6.3.

For comparison, we also conducted the Cox proportional regression analysis (Cox and Hinkley (1974)), and the results obtained are as follows,

```
## coxph(Surv(timd3, event3)~rx, data=data2)
Call: coxph(formula =Surv(timd3, event3) ~ rx, data = data2)

      coef exp(coef) se(coef)      z    p
rx -0.0867     0.917    0.163 -0.532 0.59
```

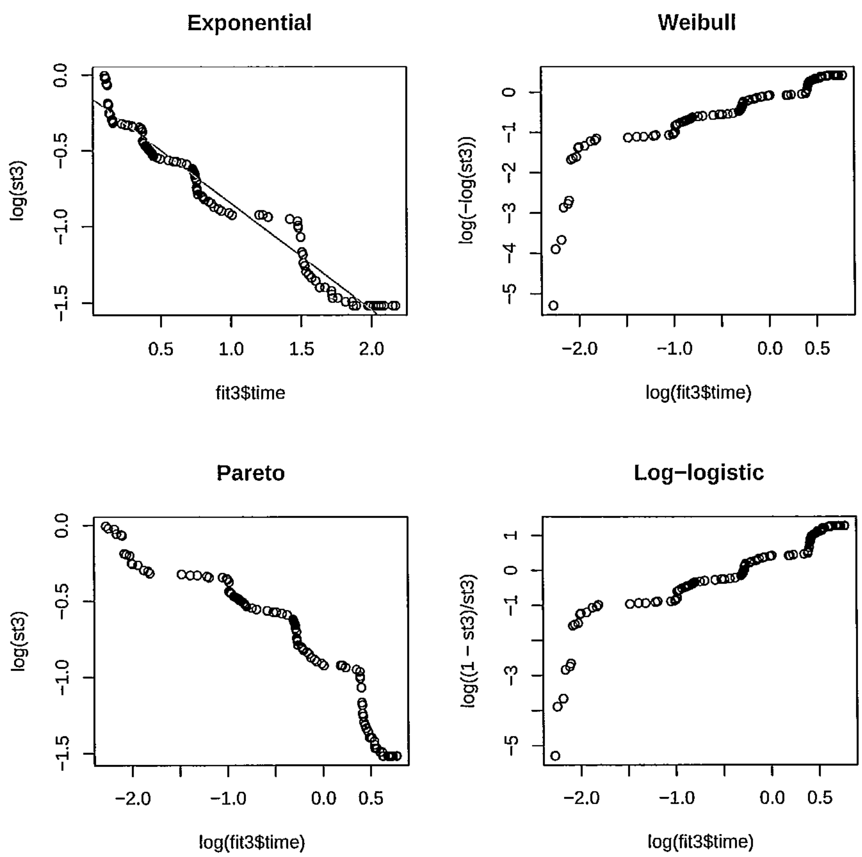

Figure 6.2: *Graphical Diagnosis.*

```
Likelihood ratio test=0.28  on 1 df, p=0.595  n=199 (4 observations
deleted due to missingness)

## coxph(Surv(timd6, event6)~rx, data=data2)
Call: coxph(formula =Surv(timd6, event6) ~ rx, data = data2)

     coef exp(coef) se(coef)     z    p
rx -0.326     0.722    0.206 -1.58 0.11
```

Table 6.3: *Estimates when loss of 3 and 6 lines distance.*

Estimation	Loss of 3 lines	Loss of 6 lines
$\hat{\alpha}_{mis}$	-0.1337	-0.8921
$\hat{\beta}_{mis}$	-0.1084	-0.3448
$\hat{\alpha}_{true}$	-0.1072	-0.8854
$\hat{\beta}_{true}$	-0.1161	-0.3483
$V_{mis}(\hat{\alpha})$	0.0130	0.0185
$V_{true}(\hat{\alpha})$	0.0133	0.0186
$V_{mis}(\hat{\beta})$	0.0265	0.0423
$V_{true}(\hat{\beta})$	0.0271	0.0425

```
Likelihood ratio test=2.53  on 1 df, p=0.112  n=199 (4 observations
## deleted due to missingness)
```

Looking at the estimates of parameter β from both methods, we found that they are close, and the tests showed the same conclusion that the treatment is not significant for the ARMD, where, we did the likelihood ratio tests for Cox PH model, and the Wald tests with respect to the likelihood estimates of the proposed parametric model, and the p-values at loss of 3 lines distance and loss of 6 lines distance were 0.78 and 0.24 respectively.

Also from Table 6.3 we found that the variances obtained from the true likelihood are greater than that from likelihood which coincides with the results from the previous previous theoretical work.

Step 3. Diagnostic checking on the assumed model

We made the histogram on the Cox-Snell residuals to see whether the proposed model suitable. From Figure 6.3 we found that the proposed model was relatively close to the theoretical one. The calculated mean of the Cox-Snell residuals was

0.78, not far from the exponential distribution with mean 1. This confirms that an exponential assumption for the data of ARMD is reasonable. Obviously, since we can not use the exact data for the calculation, there must be some difference between the proposed model and the true model.

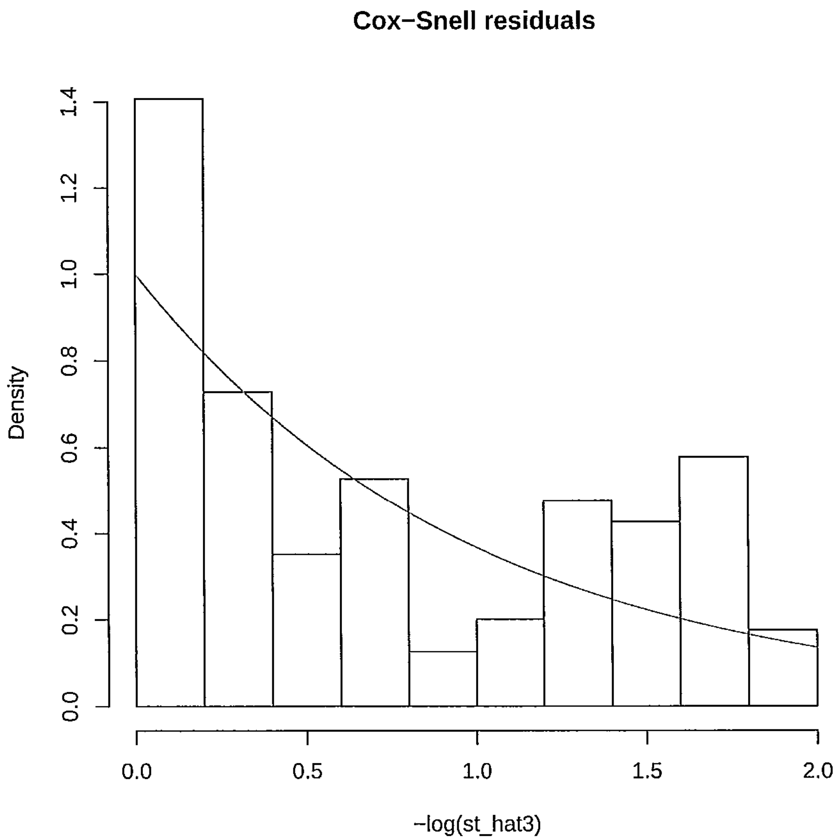

Figure 6.3: *Cox-Snell residuals.*

Chapter 7

Discussion

The work carried out in this thesis can be divided into two parts. The first part dealt with analytical comparison and numerical comparison regarding estimators from the pseudo and true likelihoods.

We have demonstrated that when the non-linear function R_i (the conditional odds on surviving in the interval) arising in the true likelihood is approximated by the first order, we get identical maximum likelihood estimates of α or λ_0 and $\boldsymbol{\beta}$ as in the pseudo-likelihood. Thus it is reasonable to conclude that the pseudo MLEs are only useful as the starting values in a numerical estimation procedures for parameters in the true likelihood. We then show that the pseudo-likelihood produces smaller variances for $\hat{\lambda}_0$ and $\hat{\boldsymbol{\beta}}$ than does the true likelihood, in the case of the exponential distribution with categorical or continuous covariates. Thus, the results obtained from the pseudo-likelihood are misleadingly better. Throughout the thesis, the end (or mid) points of the interval were used as exact times to event in the pseudo-likelihood. For the simple exponential case with $S(t) = e^{-\lambda t}$, from (4.1.2) one can show that the use of the mid point of the interval as exact time to event yields a bigger standard error of $\hat{\lambda}$; making it closer to the standard error of $\hat{\lambda}$ obtained from the true likelihood.

The second part, which implemented much of the theory developed in the first part, consisted of a simulation study which enabled a numerical comparison of the MLEs obtained using the two likelihoods. Due to time restriction, only a limited number of statistical analyses could be performed, and therefore only a partial assessment of the consequences on basing inference on the pseudo-likelihood was obtained. The difference between estimates from the two likelihoods have been analyzed by two different schedules of inspection times. We have demonstrated that there is significant evidence of difference between $E(\hat{\alpha}_{true})$ and $E(\hat{\alpha}_{mis})$ when the inspection times are frequent and regular or rare and irregular. Also $H_0 : E(\hat{\beta}_{true}) - E(\hat{\beta}_{mis}) = 0$ was rejected for the two schedules. The estimates obtained from the true likelihood with the frequent and regular schedules are more accurate than estimates from the pseudo likelihood with rare and irregular schedules.

At first sight these results are re-assuring. It seems sensible that the unwarranted assumptions of exactness about "time to event" should lead to artificial precision in the MLEs. However, it is clear that further work is required in this area and it is too early to assume that this will always be the case. Meanwhile the current findings will be of interest to the statistical community, alerting trial designers to some of the pitfalls associated with various aspects of this method of analysis, especially the use of the pseudo-likelihood.

The exponential result might lead one to anticipate that the finding would be a general property of the PH family. This can be similarly investigated by further analytical work and simulation on other PH models such as the Weibull and non-PH models.

Bibliography

Abramowitz, M. and Stegun, I. A. (1970). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* New York: Dover Publications, Ninth printing.

Barndorff-Nielsen, O. E. (1988). *Parametric Statistical Methods and Likelihood.* New York: Springer-Verlag.

Bergink, G.-J., Hoyng, C. B., Maazen, R. W. M., Vingerling, J. R., Daal, W. A. J. and Deutman, A. F. (1998). A randomised controlled clinical trial on the efficacy of radiation therapy in the control of subfoveal choroidal noevascularisation in age-related macular degeneration: radiation versus observation. *Graefe's Archive for Clinical and Experimental Ophthalmology* 236, pp. 321-325.

Blagojevic, M. (2002). Interval censored estimation methods in exponential regression model. Third Year Project. Keele University, UK.

Casella, G. and Berger, R. L. (2001). *Statistical Inference.* Duxbury, Second edition.

Collett, D. (2003). *Modelling Survival Data in Medical Research.* Chapman and Hall, Second edition.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics.* Chapman and Hall.

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society* B 30, pp. 248-275.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* 42, pp. 845-854.

Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Annales de la Société Polonaise de Mathematique, Cracovie* 6, pp. 93-116.

Hart, P. M., Chakravarthy, U., Mackenzie, G., Chisholm, I. H., Bird, A. C., Stevenson, M. R., Owens, S. L., Hall, V., Houston, R. F., McCulloch, D. W., Plowman, N. (2002). Visual outcomes in the subfoveal radiotherapy study. *Archives of Ophthalmology* 120, pp. 1029-1038.

Hillier, F. S. and Lieberman G. J. (1995). *Introduction to Operations Research.* New York: McGraw-Hill.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* John Wiley, New York.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data.* John Wiley and Sons, Second edition.

Lindsey, J. C. and Ryan, L. M. (1998). Tutorial in biostatistics methods for interval-censored data. *Statistics in Medicine* 17, pp. 219-238.

MacKenzie, G. (1999). Survival analysis for longitudinal data. *14th International Workshop on Statistical Modelling.* Graz, Austria, pp. 259-264.

Rosin, P. and Rammler, E. (1933). The laws governing the fineness of powdered coal. *Journal of the Institute of Fuel* 7, pp. 29-36.

Satten, G. (1996). Rank-based inference in the proportional hazards model for interval censored data. *Biometrika* 83, 2, pp. 355-370.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society* B 38, pp. 290-295.

Wikipedia, (2009). http://en.wikipedia.org/wiki/Weibull_distribution.

Appendix A: R Code for Exponential Distribution with one Binary Covariate

```
## irregular schedule (3,6,12,24)
nsim <-1000
nsamp<-100 #nsamp=c(100,200,500)
## control censoring rate
p<-c(0.1,0.2,0.6)
## given lambda to find true beta
lambda <- 0.1
alpha <- log(lambda)
f<-function (x,a)exp(-24*a*(exp(x)))+exp(-24*a)-p[1]
str(xmin<-uniroot(f, c(-10, 10), tol = 0.0001, a = lambda))
beta<-xmin$root
## matrix for storing the results
storemis <- matrix(nrow=nsim, ncol=4)
storetrue <- matrix(nrow=nsim, ncol=5)
## generate the covariarte
b<- rep(0:1, nsamp/2)
xbeta<- alpha+b*beta
hazard<- exp(xbeta)
## generate the time
set.seed(123456789)
for (j in 1:nsim) {
u <- runif(nsamp)
st <- (-log(1-u))/hazard
## generate the observed data
e1<- rnorm(nsamp,0,1/3)
e2<- rnorm(nsamp,0,1/3)
e3<- rnorm(nsamp,0,1/3)
e4<- rnorm(nsamp,0,1/3)
t0 <- 0
t3 <- 3+e1
```

```
t6 <- 6+e2
t12 <- 12+e3
t24 <- 24+e4
## set the initial values
stbeg <- rep(NA, nsamp)
stmid <- rep(NA,nsamp)
stend <- rep(NA,nsamp)
deltai <- rep(1,nsamp)
## get the observed values
for (i in 1:nsamp) { if (t0 <= st[i] & st[i] <= t3[i])
               {stbeg[i] <- 0
               stmid[i] <- (0+t3[i])/2
               stend[i] <- t3[i]}
               else if (t3[i] < st[i] & st[i] <= t6[i])
               {stbeg[i] <- t3[i]
               stmid[i] <- (t3[i]+t6[i])/2
               stend[i] <- t6[i]}
               else if (t6[i] < st[i] & st[i] <= t12[i])
               {stbeg[i] <- t6[i]
               stmid[i] <- (t6[i]+t12[i])/2
               stend[i] <- t12[i]}
               else if (t12[i] < st[i] & st[i] <= t24[i])
               {stbeg[i] <- t12[i]
               stmid[i] <- (t12[i]+t24[i])/2
               stend[i] <- t24[i]}
               else if (st[i] > t24[i])
               {stbeg[i] <- t24[i]
               stmid[i] <- t24[i]
               stend[i] <- t24[i]
               deltai[i] <- 0}
censp<-(nsamp-sum(deltai))/nsamp # calculate censoring rate of sample
       }
## pseudo-likelihood
```

```
mydata1<-cbind(stend,b,deltai)
expmis<-function (x,mydata)
   {
p1<-x[1]
p2 <-x[2]
x<-c(p1,p2)
              stend<-mydata1[,1]
              b<-mydata1[,2]
              deltai<-mydata1[,3]
              expp <- exp(b*p2)
              hazd <- (exp(p1))*expp
              survend <- (exp(p1))*stend*expp
logmis <- -sum(deltai*(log(hazd))-(deltai*survend)-(1-deltai)*survend)
       }
resexpmis<-nlm(expmis,c(-0.1,0.1),mydata=mydata1,hessian=TRUE,print.level=1)
fit1<-resexpmis$estimate
hess1<-resexpmis$hessian
cov1<-solve(hess1)
varerr1<-diag(cov1)
stderr1<-sqrt(diag(cov1))
storemis[j,1:2]<-fit1
storemis[j,3:4]<-varerr1
## true likelihood
mydata2<-cbind(stbeg,stmid,stend,b,deltai)
expture<-function(x,mydata)
     {
p1<-x[1]
p2<-x[2]
x<-c(p1,p2)
              stbeg<-mydata2[,1]
              stmid<-mydata2[,2]
              stend<-mydata2[,3]
              b<-mydata2[,4]
```

```
              deltai<-mydata2[,5]
              expbeg<-exp(-(exp(p1+b*p2))*stbeg)
              expend<-exp(-(exp(p1+b*p2))*stend)
              expstar<-exp(-(exp(p1+b*p2))*stmid)
              r<-expend/(expbeg)
              c1 <- r
              c2<- ifelse(c1 < 0.999, c1 ,0.999)
logtrue<--sum(deltai*(log(expbeg))+deltai*(log(1-c2))
+(1-deltai)*(log(expstar)))
    }
resexptrue<-nlm(exptrue,c(fit1[1],fit1[2]),mydata=mydata2,hessian=TRUE)
fit2<-resexptrue$estimate
hess2<-resexptrue$hessian
cov2<-solve(hess2)
varerr2<-diag(cov2)
stderr2<-sqrt(diag(cov2))
storetrue[j,1:2]<-fit2
storetrue[j,3:4]<-varerr2
storetrue[j,5]<-censp
  }
## end  main simulation loop above
## store the results
lmis<- storemis[,1]
bmis<- storemis[,2]
varlmis<- storemis[,3]
varbmis<- storemis[,4]
ltrue<- storetrue[,1]
btrue<- storetrue[,2]
varltrue<- storetrue[,3]
varbtrue<- storetrue[,4]
hatcensp<- storetrue[,5]
## calculate the bias
## pseudo
```

```
hatalpha_m<- mean(lmis)
biaslm<- hatalpha_m - alpha
hatvarl_m<- mean(varlmis)
mselm<- hatvarl_m + (biaslm^2)
hatb_m<- mean(bmis)
biasbm<- hatb_m - beta
hatvarb_m<- mean(varbmis)
msebm<- hatvarb_m + (biasbm^2)
## true
hatalpha_t<- mean(ltrue)
biaslt<- hatalpha_t - alpha
hatvarl_t<- mean(varltrue)
mselt<- hatvarl_t + (biaslt^2)
hatb_t<- mean(btrue)
biasbt<- hatb_t - beta
hatvarb_t<- mean(varbtrue)
msebt<- hatvarb_t + (biasbt^2)
## censoring rate
hatp<-mean(hatcensp)
## variance from Fisher information
fishvarl<-1/(nsamp*(1-(sum(b)/nsamp)))
fishvarb<-1/(sum(b)*(1-(sum(b)/nsamp)))
## calculate the differences between estimated variances
## and inverted fisher information
difffishl<-hatvarl_t-fishvarl
difffishb<-hatvarb_t-fishvarb
## list estimators and true values
c(hatalpha_m,hatb_m,hatalpha_t,hatb_t,alpha,beta,hatp)
## list the estimated variances and inverted Fisher information
c(hatvarl_m,hatvarb_m,hatvarl_t,hatvarb_t,
stovarl_m,stovarb_m,stovarl_t,stovarb_t,fishvarl,fishvarb)
## list the calculated mse
c(mselm,msebm,mselt,msebt)
```

```
## list the differences between estimated variances and inverted
## fisher information
c(difffishl,difffishb)
## hypothesis test on the estimators from two likelihoods
t.test(x=lmis,y=ltrue,paired=T)
t.test(x=bmis,y=btrue,paired=T)
ratiol<-log(varlmis/varltrue)
ratiob<-log(varbmis/varbtrue)
t.test(ratiol)
t.test(ratiob)
## figures of distribution
op <-par(mfrow = c(1, 2))
hist(ltrue,main="",xlab="",ylab="",sub="Distribution of")
text(-1.9,220, expression(Std.Dev==0.15),cex=0.8)
text(-1.9,210,expression(Mean==-2.29),cex=0.8)
text(-1.98,200,expression(N==1000),cex=0.8)
mtext(expression(hat(alpha)),adj=0.80,line=-29.68)
hist(btrue,main="",xlab="", ylab="",sub="Distribution of")
text(1.2,170, expression(Std.Dev==0.21),cex=0.8)
text(1.2,162,expression(Mean==0.67),cex=0.8)
text(1.13,155,expression(N==1000),cex=0.8)
mtext(expression(hat(beta)),adj=0.80,line=-29.69)
par(op)
```

Appendix B: R code for Analysis of the Real Data (ARMD) in Section 6.2

```
library(foreign)
library(survival)
data<-read.spss("ARMD_data_peng.sav")
data<-data.frame(data)
dim(data)
## get the middle point
timd3<-(data[,3]+data[,4])/2
timd6<-(data[,6]+data[,7])/2
event3<-data[,5] event6<-data[,8]
rx<-data[,9]
is.factor(rx)
rx<-as.numeric(rx)
rx<-ifelse(rx==2,1,0)
data<-cbind(data,timd3,timd6)
names(data)
data1<-cbind(timd3,event3,timd6,event6)
data1<-data.frame(data1)
timd3<-ifelse(event3==0,data[,3],timd3)
timd6<-ifelse(event6==0,data[,6],timd6)
data2<-cbind(timd3,event3,timd6,event6)
data2<-data.frame(data2)
## fit at timd3,overall K-M estimate
fit3<-survfit(Surv(timd3,event3), data=data2)
summary(fit3)
st3<-fit3$surv
plot(fit3,main="Overall Survival Functions",
ylab="Cum Survival",xlab="timd3")
## graphical checking
op<-par(mfrow=c(2,2))
plot(fit3$time,log(st3),main="Exponential")
```

```
abline(lm(log(st3)~fit3$time),col="red")
plot(log(fit3$time),log(-log(st3)), main="Weibull")
plot(log(fit3$time),log(st3), main="Pareto")
plot(log(fit3$time),log((1-st3)/st3), main="Log-logistic")
par(op)
## cox proportional regression model
coxph(Surv(timd3,
event3)~rx,data=data2)
## estimate by using the pseudo-likelihood
mydata1<-cbind(timd3,rx,event3)
mydata1<-na.omit(mydata1)
expmis <- function (x,mydata)
   {
p1<-x[1]
p2<-x[2]
x<-c(p1,p2)
              stend<-mydata1[,1]
              b<-mydata1[,2]
              deltai<-mydata1[,3]
              expp <- exp(b*p2)
              hazd<-(exp(p1))*expp
              survend<- exp(p1))*stend*expp
logmis<- -sum(deltai*(log(hazd))-(deltai*survend)-(1-deltai)*survend)
       }
resexpmis<-nlm(expmis,c(0.0,0.0),mydata=mydata1,hessian=TRUE,print.level=1)
fit1<-resexpmis$estimate
hess1<-resexpmis$hessian
cov1<-solve(hess1)
stderr1<-sqrt(diag(cov1))
varerr1<-diag(cov1)
##estimate by using the true likelihood
stbeg3<-ifelse(event3==0,data[,3],data[,4])
stend3<-ifelse(event3==0,data[,3],data[,3])
```

```
mydata2<-cbind(stbeg3,stend3,rx,event3)
mydata2<-na.omit(mydata2)
exptrue <- function (x,mydata)
     {
p1<-x[1]
p2<-x[2]
x<-c(p1,p2)
              stbeg<-mydata2[,1] stend<-mydata2[,2]
              rx<-mydata2[,3]
              event3<-mydata2[,4]
              expbeg <- exp(-(exp(p1+rx*p2))*stbeg)
              expend <- exp(-(exp(p1+rx*p2))*stend)
              expstar <- exp(-(exp(p1+rx*p2))*stend)
              r<-expend/(expbeg)
              c1 <- r
              c2 <- ifelse(c1 < 0.999, c1 ,0.999)
logtrue<- -sum(event3*(log(expbeg))+event3*(log(1-c2))
+(1-event3)*(log(expstar)))
   }
resexptrue<-nlm(exptrue,c(fit1[1],fit1[2]),mydata=mydata2,hessian=TRUE)
fit2<-resexptrue$estimate
hess2<-resexptrue$hessian
cov2<-solve(hess2)
stderr2<-sqrt(diag(cov2))
varerr2<-diag(cov2)
## results list
c(fit1[1],fit1[2],fit2[1],fit2[2],varerr1[1],varerr1[2],varerr2[1],varerr2[2])
st_hat3<-exp(-(exp(fit2[1]+(fit2[2])*(mydata2[,3])))*(mydata1[,1]))
## Cox-Snell residual graph
hist(-log(st_hat3),prob=T,main="Cox-Snell residuals")
curve(dexp(x),add=T,col="red")
mean(-log(st_hat3))
## Wald test
```

```
w1<-((fit2[2])^2)/(varerr2[2])
pvalue1<-1-pchisq(w1,df=2)
## fit at timd6 (omitted)
```