UNIVERSITY OF CALGARY

Current Status Data and Additive Hazards Regression Model

by

He Gao

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

August, 2009

UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "Current Status Data and Additive Hazards Regression Model" submitted by He Gao in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE.

Xuewen Lu

Supervisor, Dr. Xuewen Lu
Department of Mathematics and Statistics

M. D. Burke

Co-supervisor, Dr. Murray Burke
Department of Mathematics and Statistics

Badescu

Dr. Alexandru Badescu
Department of Mathematics and Statistics

A. David

Dr. Alexander David
Haskayne School of Business

August 27th, 2009

Date

Abstract

Current status data, also known as Case I interval-censored data, arise when the exact knowledge about the failure time of interest is unavailable, and it can only be seen whether failure occurred before or after a random monitoring time. In the past two decades, a number of literature on the statistical analysis of current status data has appeared. Among them, one of the most important approaches was to use the semiparametric additive hazards regression models (Lin et al., 1998) to analyze such data. In this thesis, we will give a detailed discussion on the estimation as well as inference procedures for the semiparametric additive hazards regression model involved in the analysis of current status data. Meanwhile, we propose a new method by reformulating the iterative convex minorant (ICM) algorithm (Groeneboom and Wellner, 1992) as a generalized gradient projection (GGP) (Pan, 1999) to estimate the model. The suggested approach in comparison with former methods has the advantages that it is efficient and does not involve any modeling of the monitoring times.

Acknowledgements

I would like to thank all people who have helped and inspired me during my master program study.

I especially want to thank my advisors, Prof. Murray Burke, Prof. Xuewen Lu, and Prof. Karen Kopciuk for their guidance during my research and study at University of Calgary. Their perpetual energy and enthusiasm in research had motivated all their advisees, including me. In addition, they were always accessible and willing to help their students with their research. As a result, research life became smooth and rewarding for me.

Prof. Alexandru Badescu and Prof. Alexander David deserve special thanks as my thesis committee members and advisors.

My deepest gratitude goes to my family for their unflagging love and support throughout my life; this thesis is simply impossible without them. I am indebted to my father, Zhengxing Gao, for his care and love. As a typical father in a Chinese family, he worked industriously to support the family and spare no effort to provide the best possible environment for me to grow up and attend school. He had never complained in spite of all the hardships in his life. I cannot ask for more from my mother, Suhua He, as she is simply perfect. I have no suitable word that can fully describe her everlasting love to me. I remember many sleepless nights with her accompanying me when I was suffering from bone fracture. I remember her constant support when I encountered difficulties and I remember, most of all, her sacrifices for her son. Mother, I love you. I admire my uncle, Lizhong Xu, for his talents and wisdom. He had been a role model for me to follow unconsciously and has always been one of my best counselors and friends.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Survival data, also called failure time data or time to event data, are a special type data structure for displaying numerical information. They concern the effects between certain events and their associated durations (time). Examples of the events, are often considered as the failures or survival events, which include death, the appearance of diseases or certain symptoms, the failure of a mechanical component, or the occurrences of critical events that people are observing. Initially, survival data analysis appeared in medical research (Cox, 1972 and Reid, 1994). However, due to its strong advantages on handling complex information, it had developed extensively to other disciplines. These include social-economical surveys, biological experiments, demographical investigations, epidemic monitoring and machinery examinations.

A major feature of survival data that distinguishes the analysis of them from other statistical areas, is the existence of censorship. Censored data arise when a subject's incident is known to occur only in a certain period of time. Censoring mechanisms can be quite complicated and require special methods of treatment for the analysis. Truncation is another feature that some of the survival data possess and need to be specially treated during the analysis. Two available types of truncations, known as left truncation and right truncation, have been widely discussed in the literature (Lagakos et al., 1988; Klein and Moeschberger, 2003; Gross et al., 1992). However, in this thesis, we will only focus on the analysis methods that deal with Case I interval-censored data (or current status data).

Before moving to any detailed discussions on censoring, we present an example here to give a general impression on the structure and features of survival data.

1.0.1 Lung Tumor data

In medicine, tumorigenicity experiments are used to test an agent or medicine that has been injected into the experimental organisms (usually animals) to see its controlling ability, within a specific time period, for the prevention of tumor growth inside the carriers' bodies. In this type of study, the time to tumor onset is usually of interest but not directly observable. Rather, only the death or sacrifice time of an organism is observed, and the presence or absence of a tumor at the time is known. If the occurrence of a tumor causes the immediate death of the organism, then that death moment can be considered as the exact or right-censored observation of the tumor onset time and it is appropriate to apply methods that are developed for right-censored survival data to analyze such case. However, when the occurrence of a tumor cannot cause the death of the organism in a short period of time, the time to tumor onset will thus become either less than or greater than the observed time of death or sacrifice. This is a situation where both left- and right- censored observations might exist at the same time and we name the data that follow this type of structure as current status data.

Hoel and Walberg (1972) reported the study of the tumorigenicity experiment for lung tumor; the results were reorganized by Finkelstein and Wolfe in 1985. The purpose of the experiment was to compare the onset time of lung tumor for untreated mice in a germ-free environment against a conventional environment. A total of 144 male RFM mice were selected into the study. At the times of sacrifice, 27 mice were found having lung tumors among a total of 96 mice in the conventional environment, whereas 35 cases out of 48 mice were discovered having lung tumors in the germ-free environment. The data are presented in Table 1.1.

As mentioned previously, censoring is one of the distinguishing features for survival data. In most of the circumstances, it refers to the situation where the incomplete ob-

Group	*Tumor status*	*Death times*
CE	With tumor	381 477 485 515 539 563 565 582 603 616 624 650 651 656 659 672 679 698 702 709 723 731 775 779 795 839
	No tumor	45 198 215 217 257 262 266 371 431 447 454 459 475 479 484 500 502 503 505 508 516 531 541 553 556 570 572 575 577 585 588 594 600 601 608 614 616 632 632 638 642 642 642 644 644 647 647 653 659 660 662 663 667 667 673 673 677 689 693 718 720 721 728 760 762 773 777 815 886
GE	With tumor	546 609 692 692 710 752 773 781 782 789 808 810 814 842 846 851 871 873 876 888 888 890 894 896 911 913 914 914 916 921 921 926 936 945 1008
	No tumor	412 524 647 648 695 785 814 851 880 913 942 986

Table 1.1: *Death times in days for 144 male RFM mice with lung tumors*

servation of survival time had occurred. In other words, the survival time can be known from a certain range instead of being observed exactly. Censored data should be differentiated from incomplete data as the censoring occurred contains plausible information (although it is not highly accurate) that will be required for the subsequent analysis, whereas the missing observations existed in the incomplete data have no information at all for the entire data set and instead, they need to be 'filled in' by imputation methods based on the rest of the information.

The most common type of censorship for survival data is the right censoring. Often it occurs when the failure time of interest is observed either exactly or to be greater than a censoring time. Examples that yield right-censored observations include situations when the study has to be ended due to certain restrictions such as time constraints or resource limitations. In these incidents, since some of the subjects' events have not occurred at the end of the study, their survival times are considered to be greater than the censoring time; and for the remaining subjects, because their events occurred before the end of the study, their survival times are known to be exact. One thing that needs to be paid attention for

censored survival data is the fact that the study end time may be varied in accordance with individuals' movements and a reasonable scenario is given by the withdrawal of the subjects from study due to their personal matters. Thus, in the censored survival data, there often exists a censoring variable to indicate the censoring time. For data that have been right censored, to let the observation be exact, the survival variable must be smaller than censoring variable; otherwise, the observation can be only inferred from the situation where the survival variable is greater than the censoring variable.

It is very important to know that one must fully understand the structure in which pattern the right-censoring occurs in order to analyze such data properly. To achieve this, an independence assumption needs to be imposed in prior. That is, we assume the failure rate or hazard is the same between the subjects who are still in the study and the subjects who have been censored out. More specifically, under the independence assumption, we can claim that

$$\lim_{\Delta t \to 0^+} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \lim_{\Delta t \to 0^+} \frac{P(t \leq T < t + \Delta t | T \geq t, Y(t) = 1)}{\Delta t} \tag{1.1}$$

(Kalbfleisch and Prentice, 2002), where T is the survival variable of interest, and $Y(t) = 1$ represents the corresponding subject is at risk at time t. As an alternative, the above expression is equivalent to

$$\lim_{\Delta t \to 0^+} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \lim_{\Delta t \to 0^+} \frac{P(t \leq T < t + \Delta t | T \geq t, C \geq t)}{\Delta t} \tag{1.2}$$

(Sun, 2006), where C is referred to as the censoring indicator.

Types of censorship are classified based on practical problems. For instance, the pattern in which all the subjects have the same stopping time for the study usually stands as Type I right censoring, whereas Type II right censoring often occurs when the study time stops after a fixed number of subjects have failed. In addition to right censoring, some observations may be left-censored. Left censoring means that the failure

time is known to be less than certain time. Interval Censoring, the focus of this thesis, is introduced in the later section.

In certain types of survival data, subjects can be included in the study only if their failure times meet some required conditions. These situations are named by the term, truncation. A common example that yields the truncated survival data is a cohort study in which subjects are included in the study when they experience some events prior to the survival event. This type of truncation is called left-truncation. Truncation has similar properties as censoring. For information about its discussions, readers could refer to Kalbfleisch and Prentice (2002) and Lawless (2003).

1.0.2 Survival Data with Interval Censoring

Interval censoring occurs when study subjects or interested failure times are not under continuous monitoring and the survival/failure time is thus not always exactly observed or right-censored. For data that are interval-censored, one only knows a time interval (duration), within which the event has taken place. Exact or right-censoring can be treated as a special example of interval censoring as in these incidents, the time interval converges either to a time point or diverges infinitely on the right. Further, in a more general sense, the interval-censored observations can be considered as a collection of many of these time sub-intervals or points (Turnbull, 1976).

Interval-censored survival data can be found in many scientific areas, some of which include medicine, biology, epidemiology, engineering, finance, economics, sociology, and psychology. A typical example is given by the clinical trial in medicine involved with periodic follow-ups. In this type of examples, a patient may visit the clinic at the times when they feel convenient rather than at the times that have been scheduled previously. Hence, the trial will miss out on one or more observations on the status of disease change

for that patient. At this stage, the data on that change are interval-censored. Note that in the longitudinal study, interval-censored data usually have grouped failure times, in which the observations of each subject are a member of the union of non-overlapping sub-intervals. Grouped failure time data can be analyzed in simpler ways compared with other interval-censored data and one may refer to Lawless (2003) for a detailed discussion. However, for simplicity, this thesis will not include any additional information in the subsequent texts on this subject.

To formally define the interval censoring, let T be a nonnegative random variable representing the time when a failure occurred to an individual in a survival study. An observation on T is interval-censored if a time interval $(L, R]$ is observed such that

$$T \in (L, R], \tag{1.3}$$

where $L \leq R$. If $R = \infty$, right-censoring is occurred, whereas if $L = R$, an exact observation is obtained. In this section, four common types of interval censoring will be discussed in detail.

I. Case I Interval-censored Survival Data

Case I interval-censored survival data usually refer to the situation where all the observed time intervals contain either time zeros, $L = 0$, or infinities, $R = \infty$ (Groeneboom and Wellner, 1992; Huang, 1996). In other words, the observation for each subject's failure time is either left- or right-censored. Case I interval-censored survival data occur when the failure time of interest cannot directly be observed, rather it can be known only by the indication as whether it's located below or above a random monitoring time. For convenience, here a simpler notation is used to describe Case I interval-censored survival data. That is, $\{C, \delta = I(T \leq C)\}$, where C denotes a random monitoring time and I is an indicator function. Note that Case I interval-censored survival data are different from right-censored data or left-censored data, since the latter case contains exact observations

but the former one does not.

Case I interval-censored survival data are also called current status data, a term which originated from demographical studies. Cross-sectional studies and tumorigenicity experiments are two common areas that frequently generate Case I interval-censored survival data. Note that there is a fundamental difference for the current status data generated from these two perspectives. Often the data obtained from the former is dependent on the study design whereas those produced by the latter usually results from the failure of measurement accuracy of the variables.

II. Case II Interval-censored Survival Data

Case II interval-censored data are yielded when interval-censored data include at least one interval $(L, R]$ with both L and R belonging to $(0, \infty)$ (Groeneboom and Wellner, 1992; Huang and Wellner, 1997; Sun, 1998, 2005). In other words, Case II interval-censored data are the one in which some of the finite time intervals contain no zeros. A convenient notation to represent a Case II interval-censored survival data is

$$\{U, V, \delta_1 = I(T \leq U), \delta_2 = I(U < T \leq V), \delta_3 = 1 - \delta_1 - \delta_2\} \tag{1.4}$$

given the assumption that each subject is observed twice, with each of which corresponded to time U and V respectively, to satisfy the condition of $P(U \leq V) = 1$. This formulation is often used in the theoretical investigation of inference procedure. Note that once letting $U = V = C$, (1.4) can be used to describe the current status data. In the literature, Case II interval-censored survival data are sometimes just called interval-censored data, which are to underlie their general properties among other types.

Another way to present Case II interval-censored survival data is to assume a group of observed time points U_i having a order constraint, $U_1 \leq U_2 ... \leq U_K$ where K is a random integer, for each subject. Then, the above statement becomes

$$\{(K, U_j, \delta_j = I(U_{j-1} < T < U_j)), j = 1, ..., K\}, \tag{1.5}$$

where $U_0 = 0$. This is often referred to as case K or mixed case interval-censored data (Schick and Yu, 2000; Wellner, 1995). It can be seen that the formulations of (1.4) and (1.5) offer the interval-censored survival data a natural presentation to describe certain types of studies often involved with periodic follow-ups.

All three expressions, (1.3) to (1.5), will eventually lead the likelihood function to have a similar form. Meanwhile, aside from the sensible structure and ease of understanding of models (1.4) and (1.5), one usually needs to impose assumptions such as independence with T on them. Given any data following (1.4) or (1.5), one can easily transform the corrsponding model into the form of (1.3). However, the reverse is currently difficult to achieve with the absence of extra information, and special treatments are thus being called for to resolve the issue.

III. Doubly Censored Survival Data

Suppose that a survival study involves two related events and let X and S be the corresponding times for these two events. Define T to be the survival time of interest with the relationship, $T = S - X$. Thus, a doubly censoring occurs where in place of knowing X and S exactly, one only observes two intervals $(L, R]$ and $(U, V]$ so that

$$X \in (L, R], \ S \in (U, V]$$

subject to the conditions $L \leq R$ and $U \leq V$. This is different from the types of interval-censoring that so far have been discussed. However, it can be seen that a double-censoring is formed when the observations on both X and S are interval-censored.

Consider a special type of doubly censored data where S is only right-censored and in this case, one has either $U = V$ or $V = \infty$. An alternative description for this special type can be expressed by involving a censoring variable C, which is often assumed to be independent of S, so that the observation on S consists of $S^* = \min\{S, C\}$ and $\delta = I(S^* = S)$, where I is the indicator function as defined before.

IV. Panel Count Data

In studies of recurrent events, a panel count data are yielded when each subject is observed only at finite discrete time points instead of continuous time points (Sun and Wei, 2000; Kalbfleisch and Lawless, 1985). In those settings, one only knows the number of occurrences of the events between observation times and no information however is available on subjects between the observation time points. In addition to the term, panel count data, others also name it as interval count data or interval-censored recurrent data (Lawless and Zhan, 1998; Thall, 1988) in the sense where, if the event can occur only once, then the data become interval-censored survival data.

Panel count data arise in many applications, some of which include demographical and biological investigations. A typical example which can demonstrate its form is the cancer follow-up study. In these types of studies, one would be interested in the recurrent rate of various types of tumors or of these tumors at various locations. However, due to the difficulties of following the subjects continuously, all that is known is the information about the number of occurrences of the events between observation times, thus a panel data set is obtained. Another example is economic surveys in the labour market.

V. Independence for Interval Censoring and Miscellaneous Remarks

In terms of independence for interval-censoring, we mean that the censoring pattern that follows a structure of interval-censored survival data is independent of the variables of interest. For current status data, this implies that C and T are independent. For interval-censored data described by the expressions (1.4) and (1.5), this means the joint likelihood function of U and V or U_i's has no information at all to the parameters that involve with the event time T. Given a data set that follows the general expression (1.3), the independence assumption for interval-censoring assumes that the interval $(L, R]$ should include all the information that the event time T has possessed. Transforming it

into mathematical expression, this can be formulated as

$$P(T \leq t | L = l, R = r, L \leq T < R) = P(T \leq t | l \leq T < r)$$

(Self and Grossman, 1986; Zhang et al., 2005). Under the setting of independence, one does not have to be concerned with the censoring pattern in analyzing an interval-censored data set. Throughout the thesis, independence for interval-censoring is assumed unless otherwise specified.

To express an interval-censored observation, one could also use $[L, R]$, $[L, R)$, or (L, R) (Peto, 1973; Turnbull, 1976). If T is continuous, there will be no dissimilarities among these expressions as they represent the same information being observed about T; however, if T is discrete, an inconsistency can occur because the information they contain about T is different from each other. Ng (2002) discussed this circumstance in a great detail. But in this thesis, we will use the notation $(L, R]$ as the standard expression for the rest of the chapters. As mentioned earlier, the exact and right-censored observations can be treated as a special example of the interval-censored data. Suppose that T is continuous, then for an exact observation $T = t_0$, its likelihood function is $f(t_0)$. And for an interval-censored observation $(L, R]$, the likelihood function has the form $S(L) - S(R)$. Here $f(t)$ and $S(t) = P(T > t)$ represent the density function and survival function for T, respectively.

Throughout the remaining chapters, we will focus our attentions on Case I interval-censored survival data, known as the current status data. The main objective of this thesis is to demonstrate the estimation and inference procedures for the additive hazards regression model applied to such data. To achieve this, the article is organized as the following. In Chapter 2, we will discuss the concepts of counting processes with some basic quantities that are commonly used in survival analysis. Chapter 3 will give an introduction to Nonparametric Maximum Likelihood Estimator (NPMLE) and illustrate

the ICM algorithm which is designed to provide the empirical results for nonparametric estimators. Chapter 4 will talk about the regression models, particularly with a special emphasis on the analysis of the additive hazards regression. Simulation studies and a real example demonstration will be presented in Chapter 5. To end this thesis, conclusions including extra knowledge remarks, current issues as well as comments on future work will be provided in Chapter 6.

Chapter 2

Basic Quantities and Counting Processes

2.0.3 Basic Quantities for Survival Data

In this section, some basic identities that are often used in survival analysis will be discussed. The content of the discussion will include the definition of these identities as well as their interrelationships. However, to keep our demonstration simple at this stage, we don't consider any effects generated by covariates. Let T be the time to the occurrence of some specific events. These events, on one hand, might include death, appearance of disease, machine breakdown; and on the other hand, may refer to remission of some treatments, conception, cessation of smoking and so forth. More specifically, in this section, T is a nonnegative random variable from a homogeneous population. Three functions can be used to describe the distribution of T, namely, the survival function, which is the probability of a subject 'surviving' beyond time T; the hazard rate function, also called risk function, which describes the conditional chance a subject at time t experiences the event in the next instant and the probability density function, as referred to be the probability of the event's occurring at time t. If any one of these three identities is known, the other two can be explicitly determined. In practice, these three functions, together with another important quantity, the cumulative hazard function are used to illustrate different aspects of the distribution of T in survival analysis. In the later chapters of this thesis, we will see how these functions are estimated and how inferences are developed based on their estimation.

The survival function is defined as

$$S(t) = P(T > t), \tag{2.1}$$

which describes the likelihood of which an subject survives after an event occurs at some specific time t. If T is continuous, then $S(t)$ is a continuous and strictly decreasing function. Moreover, the survival function becomes the complement of the cumulative distribution function, namely, $S(t) = 1 - F(t)$ so that $F(t) = P(T \leq t)$. In general, the above expression can be formulated as

$$S(T) = P(T > t) = \int_t^{\infty} f(x)dx, \tag{2.2}$$

where

$$f(t) = -\frac{dS(t)}{dt}.$$

Note that $f(t)$ is the probability distribution function with the properties of being non-negative and of being equal to one after integrating it from $-\infty$ to $+\infty$. When T is discrete, say its probability mass function $P(t_j) = P(T = t_j), j = 1, 2, ...,$ counts on values t_j where $j = 1, 2, ...$ and $t_1 < t_2 < ...$, then the survival function of T will be defined by

$$S(t) = P(T > t) = \sum_{t_j > t} p(t_j). \tag{2.3}$$

Despite of various types of survival function that might present in survival analysis, they all share common characteristics with each other and have same properties. Specifically, they are monotone non-increasing functions having values between 0 and 1 as time progresses. Their rate of decline varies upon the risk of experiencing the event at time t and can be very useful in comparing multiple failure patterns.

Another quantity that is fundamental and crucial in survival analysis is the hazard function. This function is also known as the intensity function in stochastic processes, or simply as the hazard rate. The hazard function is defined by

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P[t \leq T < t + \Delta t | T \geq t]}{\Delta t}. \tag{2.4}$$

When T is continuous,

$$\lambda(t) = f(t)/S(t) = -dln[S(t)]/dt. \tag{2.5}$$

In terms of the cumulative hazard function $\Lambda(t)$, it has the form

$$\Lambda(t) = \int_0^t \lambda(u)du = -ln[S(t)]. \tag{2.6}$$

Hence, for the continuous case,

$$S(t) = \exp[-\Lambda(t)] = \exp[-\int_0^t \lambda(u)du]. \tag{2.7}$$

It can be seen from (2.4) that the chance for a subject at time t experiencing the event in the next instant will likely be as $\lambda(t)\Delta t$. This function is very useful in determining the failure distributions describing the information about the pattern of failure and the way in which the likelihood of experiencing the event changes over time. Like the survival function, the hazard rate, $\lambda(t)$ must be nonnegative as well.

When T is discrete, the hazard function is given by

$$\lambda(t_j) = P(T = t_j | T \geq t_j) = \frac{p(t_j)}{S(t_{j-1})}, j = 1, 2, ... \tag{2.8}$$

Since the survival function, in the discrete case, can be written as

$$S(t) = \prod_{t_j \leq t} S(t_j)/S(t_{j-1}). \tag{2.9}$$

Therefore, it can be expressed in terms of hazard function by

$$S(t) = \prod_{t_j \leq t} [1 - \lambda(t_j)]. \tag{2.10}$$

The interrelationships of these identities, for a continuous case, is summarized as the

following

$$\begin{aligned} S(t) &= \int_t^{\infty} f(x)dx \\ &= \exp[-\int_0^t \lambda(u)du] \\ &= \exp[-\Lambda(t)], \end{aligned}$$

$$\begin{aligned} f(t) &= -\frac{d}{dt}S(t) \\ &= \lambda(t)S(t), \end{aligned}$$

$$\begin{aligned} \lambda(t) &= -\frac{d}{dt}ln[S(t)] \\ &= \frac{f(t)}{S(t)}. \end{aligned}$$

2.0.4 Counting Processes

Counting process is an important approach to develop the inference procedures for censored and truncated data. This methodology was first introduced by Aalen (1975). He combined components of continuous time martingale theory, stochastic process and counting process theory into an area which allows survival quantities to build their inference properties based on the censored and truncated data. Although the complete explanation of this theory is beyond the scope of this thesis and in this section, we will only give a brief illustration; for more detailed information in this regard, one could consult books by Andersen et al. (1993), Fleming and Harrington (1991) and Klein and Moeschberger (2003).

We begin with our demonstration by defining the counting process as the following.

Suppose a stochastic process, $N(t), t \geq 0$, is a counting process if it has the property that $N(0) = 0$; $N(t) < \infty$ with probability one; and the sample paths of $N(t)$ are right-continuous and piecewise constant with jumps of size one. Given a current status example, the counting processes are, $N_i(t) = I[C_i \leq \min(T_i, t)], i = 1, 2, \ldots n$, which jumps by one at time t when $C_i = t$ and $T_i \geq t$. Here, the censoring indicator under this setting is defined by $\delta_i = I(C_i \geq T_i)$. The sum of the individual counting processes, $N(t) = \sum_{i=1}^{n} N_i(t)$ is also a counting process and it simply counts the number of failures in the sample at or prior to time t.

The counting process shows us the occurrence time of the event. Besides knowing of it, we also have information about the subjects at time t. For example, given a current status data, this information at time t contains message of who failed at or prior to time t and who has been censored prior to time t. The sum of this information that presents the occurrence of the events to subjects up to time t is called the history of filtration of the counting process at time t and is denoted by $\mathcal{F}$. As time moves on, an increasing knowledge from the sample can be learned so that $\mathcal{F}_s \subset \mathcal{F}_t$ will be obtained for $s \leq t$. In the case of current status data, the history at time t, $\mathcal{F}_t$, consists of information about one group of subjects whose $C_i \leq t$ and $C_i > t$ for the other group of subjects that are still under the study at time t. Let's denote the history just prior to time t by $\mathcal{F}_{t-}$. The history $\mathcal{F}_t, t \geq 0$ for a given problem depends on the information available up to and including time t.

For current status data, given the independence assumption among the monitoring times, C_is, the likelihood of an event at time t given the history prior to time t has the

form (Lin et al., 1998),

$$P(t \leq C_i \leq t + dt, \delta_i = 0 | \mathcal{F}_{t-}) = \begin{cases} P(t \leq C_i \leq t + dt, T_i \geq C_i | C_i > t) = h(t)\, dt & \text{if } C_i \geq t \\ 0 & \text{if } C_i < t \end{cases} \tag{2.11}$$

where $h(t)dt = dH_0(t) = e^{-\Lambda_0(t)} d\Lambda_c(t)$ is a hazard rate function (details are shown in Chapter 4).

Let's define $dN(t) = N[(t + dt)^-] - N(t^-)$ (Here t^- is a time just prior to time t) to be the change in the process $N(t)$ over a short time interval $[t, t + dt)$. Given the current status data (assume no ties are present), $dN(t)$ is one if $C_i = t$ and a subject has been failure-free up to t or 0, otherwise. If we define the process $Y(t)$ as the number of subjects i such that $C_i \geq t, 1 \leq i \leq n$, then by (2.11),

$$\begin{aligned} E[\, dN(t) | \mathcal{F}_{t-}] &= E[\text{Number of observations with} \\ &\quad t \leq C_i \leq t + dt, T_i \geq C_i | \mathcal{F}_{t-}] \\ &= Y(t) h(t)\, dt \end{aligned}$$

(Klein and Moeschberger, 2003). The process $a(t) = Y(t)h(t)$ is called the intensity process of the counting process. Further, it is a stochastic process that relies on the information included in the history, $\mathcal{F}_t$, through $Y(t)$.

The cumulative intensity process, $A(t)$ is defined by $\int_0^t a(s)\, ds, t \geq 0$. This process has the property that $E[N(t)|\mathcal{F}_{t-}] = E[A(t)|\mathcal{F}_{t-}] = A(t)$. The stochastic process $M(t) = N(t) - A(t)$ is called the counting process martingale. This process has the property that, given the strict history, F_{t-}, the future increments of this process have zero expected

value. To verify this,

$$\begin{aligned} E(\,dM(t)|\mathcal{F}_{t-}) &= E[\,dN(t) - dA(t)|\mathcal{F}_{t-}] \\ &= E[\,dN(t)|\mathcal{F}_{t-}] - E[a(t)\,dt|\mathcal{F}_{t-}] \\ &= 0 \end{aligned}$$

(Anderson et al., 1993; Fleming and Harrington, 1991). Since $a(t)$ has a fixed value given $\mathcal{F}_{t-}$, the last equality holds.

In general, a stochastic process can be called a martingale if such process is integrable and adaptable and its expected value at time t, given its past at time $s < t$, has the same value as when it is at time s. That is, $M(t)$ is a martingale if

$$E[M(t)|\mathcal{F}_s] = M(s), \text{for all } s < t. \tag{2.12}$$

To see this quantity is equivalent to having $E[\,dM(t)|\mathcal{F}_{t-}] = 0$ for all t, note that, if $E[\,dM(t)|\mathcal{F}_{t-}] = 0$, then,

$$\begin{aligned} E[M(t)|\mathcal{F}_s] - M(s) &= E[M(t) - M(s)|\mathcal{F}_s] \\ &= E[\int_s^t dM(u)|\mathcal{F}_s] \\ &= \int_s^t E[E[\,dM(u)|\mathcal{F}_{u-}]|\mathcal{F}_s] \\ &= 0 \end{aligned}$$

(Klein and Moeschberger, 2003).

The counting process martingale, $M(t) = N(t) - A(t)$ is formed by two parts. The first part is the process $N(t)$, which is a non-decreasing step function. The second part $A(t)$ is a smooth process where its value is fixed at time t. This function is called a compensator of the counting process. The martingale has a zero mean, since $E(M(t)) = E(E[M(t)|\mathcal{F}_0]) = E(M(0)) = 0$.

Another important identity needed for the counting process theory is a quantity called the predictable variation process of $M(t)$, symbolized by $\langle M\rangle(t)$. This quantity acts like a compensator of the process $M^2(t)$ and the name of it, predictable variation process, comes from the fact that $\text{var}(\,dM(t)|\mathcal{F}_{t-}) = d\langle M\rangle(t)$ for a martingale $M(t)$. To see this, we use the property, $E[\,dM(t)] = 0$. Now,

$$\begin{aligned} dM^2(t) &= M[(t+\,dt)^-]^2 - M(t^-)^2 \\ &= [M(t^-) + \,dM(t)]^2 - M(t^-)^2 \\ &= [\,dM(t)]^2 + 2M(t^-)\,dM(t). \end{aligned}$$

(Klein and Moeschberger, 2003).

Therefore,

$$\begin{aligned} \text{Var}[\,dM^2(t)|\mathcal{F}_{t-}] &= E[(\,dM(t))^2|\mathcal{F}_{t-}] \\ &= E[(\,dM(t))|\mathcal{F}_{t-}] - 2E[M(t^-)\,dM(t)|\mathcal{F}_{t-}] \\ &= d\langle M\rangle(t) - 2M(t^-)E[\,dM(t)|\mathcal{F}_{t-}] \\ &= d\langle M\rangle(t) \end{aligned}$$

(Klein and Moeschberger, 2003) and $M(t^-)$ is a fixed number as $E[\,dM(t)|\mathcal{F}_{t-}] = 0$. For $\text{Var}[\,dM(t)|\mathcal{F}_{t-}]$, it can be shown that $\text{Var}[\,dM(t)|\mathcal{F}_{t-}] \cong a(t) = Y(t)h(t)$ given the scenario when no ties will be present in the censored data.

In counting process theory, $K(t)$ is usually denoted to be a predictable process, which is considered as a stochastic process that can be known, given the history just prior to time t, $\mathcal{F}_{t-}$. An example of it is the process $Y(t)$. Over the interval 0 to t, the stochastic integral for this type of process, with respect to a martingale, is defined by $\int_0^t K(u)dM(u)$. Such stochastic integrals themselves are martingales in terms of t and their predictable variation process can be found from the identity,

$$\langle \int_0^t K(u)\;dM(u)\rangle = \int_0^t K(u)^2 d\langle M\rangle(u) \tag{2.13}$$

(Anderson et al., 1993; Fleming and Harrington, 1991; Klein and Moeschberger, 2003).

These quantities can be used to estimate the cumulative hazard rate $H(t)$ based on current status data. Recall that $dN(t) = Y(t)h(t)\,dt + dM(t)$. Then, according to Klein and Moeschberger (2003), if $Y(t)$ is nonzero,

$$\frac{dN(t)}{Y(t)} = h(t)\,dt + \frac{dMt}{Y(t)}. \tag{2.14}$$

And the expectation of it, given the fixed value of $Y(t)$ prior to time t, will be

$$E[\frac{dM(t)}{Y(t)}|\mathcal{F}_{t-}] = \frac{E[\,dM(t)|\mathcal{F}_{t-}]}{Y(t)} = 0.$$

In the mean time, the conditional variance has the form

$$\mathrm{Var}[\frac{dM(t)}{Y(t)}|\mathcal{F}_{t-}] = \frac{\mathrm{Var}[\,dM(t)|\mathcal{F}_{t-}]}{Y(t)} = \frac{d\langle M\rangle(t)}{Y(t)^2}.$$

Suppose a $J(t)$ is defined to indicate whether $Y(t)$ is positive and let $0/0 = 0$, then, by integrating both sides of equation (2.14), we can obtain an identity, of which

$$\int_0^t \frac{J(u)}{Y(u)}\,dN(u) = \int_0^t J(u)h(u)\,du + \int_0^t \frac{J(u)}{Y(u)}\,dM(u).$$

The integral $\int_0^t \frac{J(u)}{Y(u)}\,dN(u) = \hat{H}(t)$ is the Nelson-Aalen type estimator of $H(t)$. The stochastic integral, $W(t) = \int_0^t \frac{J(u)}{Y(u)}\,dM(u)$, is the predictable process $\frac{J(u)}{Y(u)}$ with respect to a martingale (Note that the integral itself is also an martingale). Again, this integral can be considered as random noise in our estimate. The random quantity $H^*(t) = \int_0^t J(u)h(u)\,du$, for current status data is equal to $H(t)$ under the condition that by ignoring the random noise in $W(t)$ in the present data, the statistic $\hat{H}(t)$ is a nonparametric estimator of the identity $H^*(t)$.

Note that the expectation of $H^*(t)$ is equal to $H(t)$. The predictable variation process

of $W(t)$ can be derived, using equation (13), as

$$\begin{aligned}\langle W\rangle(t) &= \int_0^t [\frac{J(u)}{Y(u)}]^2\, d\langle M\rangle(u) \\ &= \int_0^t [\frac{J(u)}{Y(u)}]^2 Y(u)h(u)\, du \\ &= \int_0^t [\frac{J(u)}{Y(u)}]h(u)\, du\end{aligned}$$

(Anderson et al., 1993; Fleming and Harrington, 1991; Klein and Moeschberger, 2003).

The last definition that needs to be discussed here for the counting process approach is the martingale central limit theorem. To illustrate, we treat $Y(t)/n$ and $N(t)/n$ to be sample averages for a given data set and if the sample size is large enough, the variation in both terms should be small. For large n, suppose that $Y(t)/n$ is close to a function $y(t)$, which can be simply determined. Let $Z^n(t) = \sqrt{n}W(t) = \sqrt{n}[\hat{H}(t) - H^*(t)]$. This process is nearly equal to $\sqrt{n}[\hat{H}(t) - H(t)]$. Thus, based on the historical information that has been given, the variance of the changes in $Z^n(t)$ will converge to $h(t)/y(t)$. To check this

$$\begin{aligned}\mathrm{Var}[\,dZ^n(t)|\mathcal{F}_{t-}] &= n\mathrm{Var}[\,dW(t)|\mathcal{F}_{t-}] \\ &= n\mathrm{Var}[\frac{dM(t)}{Y(t)}|\mathcal{F}_{t-}] \\ &= n\frac{d\langle M\rangle(t)}{Y(t)^2} \\ &= n\frac{a(t)\,dt}{Y(t)^2} \\ &= n\frac{Y(t)h(t)\,dt}{Y(t)^2} = \frac{h(t)\,dt}{Y(t)/n}\end{aligned}$$

(Anderson et al., 1993; Fleming and Harrington, 1991; Klein and Moeschberger, 2003), which converges to $h(t)\,dt/y(t)$ for large n. Also, Z^n will have jumps of order $1/\sqrt{n}$, which are almost continuous and a predictable variation process of it will be approximately equal to

$$\langle Z^n\rangle \approx \int_0^t \frac{h(u)\,du}{y(u)}. \tag{2.15}$$

It turns out that the limiting process, Z^{∞} is unique given the condition of having it a martingale with continuous sample paths and of which its predictable variation $\langle Z^{\infty} \rangle$ is exactly equal to (2.15). This limiting process follows finite-dimensional normal distributions along with independent increments. According to (Klein and Moeschberger, 2003), a process has independent increments if, for any set of nonoverlapping intervals $(t_{i-1}, t_i), i = 1, ..., k$ the random variables $Z^{\infty}(t_i) - Z^{\infty}(t_{i-1})$ are independent, whereas the limiting process follows finite-dimensional normal distributions if the joint distribution of $[Z^{\infty}(t_i), ..., Z^{\infty}(t_k)]$ is multivariate normal for any value of k. For the process $Z^{\infty}, [Z^{\infty}(t_1, ..., Z^{\infty}(t_k)]$ has a k-variate normal distribution with mean 0 and a covariance matrix with elements

$$\text{cov}[Z^{\infty}(t), Z^{\infty}(s)] = \int_0^{\min(s,t)} \frac{h(u)\, du}{y(u)}$$

(Klein and Moescherberger, 2003).

Moreover, $\sqrt{n}[\hat{H}(t) - H^*(t)]$ will approximately normally distributed with mean 0 and variance

$$\sigma[Z^{\infty}] = \int_0^t \frac{h(u)\, du}{y(u)},$$

which can be replaced by the form

$$n \int_0^t \frac{dN(u)}{Y(u)^2}$$

due to the fact that we can estimate $y(t)$ by $Y(t)/n$ and $h(t)$ by $dN(t)/Y(t)$. Thus, the convergence will enable us to find confidence intervals for the cumulative hazard rate at a specific time.

Counting processes can be also used to construct the likelihoods for survival data. To derive a likelihood function, let's consider each individual process, $N_i(t)$, in the study. Based on the historical information which is known up to time t, $dN_i(t)$ follows approximately a Bernoulli distribution with the corresponding probability, $P(\, dN_i(t) = 1) =$

$a_i(t)\,dt$. The function of likelihood at a given time is, then, proportional to

$$a_i(t)^{\,dN_i(t)}[1 - a_i(t)\,dt]^{1-\,dN_i(t)}$$

(Anderson et al., 1993; Fleming and Harrington, 1991; Klein and Moescherberger, 2003).

For current status data, where $a_i(t) = Y_i(t)h(t)$, with $Y_i(t) = 1$ if $t \leq C_i$; 0 if $t > C_i$, so

$$L \propto \prod_{i=1}^{n}[\exp(-H(C_i))^{\delta_i} \cdot (1 - \exp(H(C_i)))^{1-\delta_i}].$$

Chapter 3

Nonparametric Maximum Likelihood Estimation for Current Status Data

Estimating survival function is perhaps the most commonly needed task for the analysis of survival data. There can be many reasons for such a task and an example among these reasons might be that an assumption of a particular regression model requires to be evaluated by an estimated survival function for the underlying survival variable of interest. Additionally, one may also need to estimate survival functions to calculate certain survival probabilities, to graphically perform the comparison of different treatments, or to predict the chance of survival for future patients. In the situation where a parametric model is assumed for the underlying survival function, the estimation process is relatively simple and an maximum likelihood estimation approach can be used for such case; yet when semi-parametric models are present, the conventional method will not be appropriate for the analysis and a solution for this instant is needed. In this chapter, we will introduce a new estimation approach, namely, nonparametric maximum likelihood estimation, to estimate the distribution function of semi-parametric regression models.

For a right-censored survival data, the nonparametric maximum likelihood estimator (NPMLE) of a distribution function is the well-known Kaplan-Meier estimator (Kaplan and Meier, 1958; Kalbfleisch and Prentice, 2002). It is formed in a product-limit way and has been studied extensively for several decades. Its variance estimate becomes available after the Greenwood's formula (Greenwood, 1926) has been used. Unlike parametric inference, in the case of interval-censored survival data, nonparametric inference is more complicated than that for the right-censored data from both practical and theoretical

points of views. Specifically, the NPMLE of a distribution function does not have a closed form in general and can only be determined by using an iterative algorithm.

At first, we will discuss the derivations for the NPMLE of distribution functions based on Case I interval-censored or current status data. For this special type of interval-censored data, a closed form is available for the non-parametric maximum likelihood estimator. Then, an illustration about the algorithms that are used for determining the NPMLE applied to the same data structure will be followed.

3.0.5 Nonparametric Maximum Likelihood Estimator for Current Status Data

Let T_i's be the survival time of interest with distribution function $F(t)$. We suppose that the observations in such type data take the form

$$\{C_i, \delta_i = I(T_i \leq C_i)\},$$

where $i = 1, ..., n$ for a finite sample size. Here, C_i represents the observation time for an individual i independent of T_i and $\delta_i = I(T_i \leq C_i)$ is the indicator function for censoring status. Then, based on this information, the likelihood function can be constructed as

$$L(F(t)) = \prod_{i=1}^{n} [1 - F(C_i)]^{1-\delta_i} [F(C_i)]^{\delta_i}.$$

For the log-likelihood function, it has the form

$$l(F(t)) = \sum_{i=1}^{n} [(1 - \delta_i)\log(1 - F(C_i)) + \delta_i \log(F(C_i))]. \tag{3.1}$$

Let $C'_{(j)}$ where $j = 0, ..., r$ be the j-th order statistics of $\{0, C_i; i = 1, ..., n\}$. Define $a_j = \sum_{i=1}^{n} \delta_i I(C_i = C'_{(j)})$, the number of failed subjects observed at $C'_{(j)}$, and $n_j = \sum_{i=1}^{n} I(C_i = C'_{(j)})$, the number of subjects observed at $C'_{(j)}, j = 1, ..., r$. Then the log-

likelihood function $l(F(t))$ can be re-expressed by

$$\begin{aligned} l(F(t)) &= \sum_{j=1}^{r}[(n_j - a_j)\log(S(C'_{(j)})) + (a_j)\log(1 - S(C'_{(j)})) \\ &= \sum_{j=1}^{r}[(a_j)\log(F(C'_{(j)})) + (n_j - a_j)\log(1 - F(C'_{(j)}))], \end{aligned} \tag{3.2}$$

which can be considered as the log-likelihood is formed upon an r-dimensional binomial sample with $F(t) = 1 - S(t)$.

The log-likelihood function l provides the estimation of S or F only through its values at the $C'_{(j)}$'s. Therefore, we can show that maximizing $l(F(t))$ with respect to $F(C'_{(j)})$ is equivalent to minimizing

$$\sum_{j=1}^{r} n_j[\frac{a_j}{n_j} - F(C'_{(j)})]^2$$

subject to $F(C'_{(1)}) \leq ... \leq F(C'_{(r)})$ (Robertson et al., 1988). The set of values of $C'_{(j)}$ that minimize this summation is commonly referred to as the results solved from the least squares problem for isotonic regression of $\{a_1/n_1, ..., a_r/n_r\}$ with weights $\{n_1, ..., n_r\}$ (Barlow et al., 1972; Robertson et al., 1988). By using the max-min formula for isotonic regression, the NPMLE of F at time $C'_{(j)}$ can be derived as

$$\hat{F}(C'_{(j)}) = \max_{s \leq (j)} \min_{r \geq (j)} \frac{\sum_{j=s}^{r} a_j}{\sum_{j=s}^{r} n_j}.$$

Hence, the NPMLE of F has a closed form. To compute $\hat{F}(C'_{(j)})$, one could adopt some algorithms that are often used to calculate the isotonic regression, e.g. the pool adjacent violators algorithm (PAVA).

3.0.6 Iterative Convex Minorant Algorithm for Current Status Data

For the past two decades, there has been an increasing research interest in the development of computational algorithm for determining the NPMLE of the survival

function or the cumulative distribution function. Among these algorithms, three of the well-known ones include, the Self-Consistency algorithm that was developed by Turnbull (1976) and can be considered as a specific application of the EM algorithm; the Iterative Convex Minorant (ICM) algorithm written by Groeneboom and Wellner (1992), and a hybrid EM-ICM algorithm proposed by Wellner and Zhan (1997), which combines both the self-consistency algorithm and the ICM algorithm as a whole. Since we used the ICM algorithm to implement the estimation process for the additive hazards regression model, in this section, detailed discussions will only be given for that algorithm. For the rest of them, readers can find their descriptions from the sources that have been mentioned above.

To illustrate the ICM algorithm, first let

$$W = \{F = (F_1, ..., F_{r-1})' \in \Re^{r-1}; 0 \leq F_1 \leq ... \leq F_{r-1} \leq 1\},$$

be a subspace of $\Re^{r-1}$, and define $F(C'_{(j)}), j = 1, ..., r$ as the corresponding distribution function for the distinct order statistic $C'_{(j)}$ such that $F(C'_{(0)}) = 0, F(C'_{(r)}) = 1$ and $F_{C'} = (F(C'_{(1)}), ..., F(C'_{(r-1)}))'$. Then, the NPMLE of the distribution function can be obtained by maximizing the log-likelihood function $l(F(t))$ in (3.2) over W.

The realization of the ICM algorithm relies on the following two conditions. First, suppose that g and W are a differentiable concave function mapped from $\Re^{r-1}$ to $\Re$ and a convex cone in $\Re^{r-1}$, respectively. We let Q be a positive definite $(r-1) \times (r-1)$ matrix with a fixed point y in $\Re^{r-1}$ and assume that $g(F)$ attains its maximum over region W at $\hat{F}$. Define

$$g^*(F|y, Q) = -\frac{1}{2}(F - y)'Q(F - y)$$

for $F \in \Re^{r-1}$ and suppose that $\hat{F}^* \in W$ maximizes $g^*(F|y, Q)$ over W. Then, $g^*(F|y, Q)$ can be maximized by $\hat{F}^*$ over W if and only if $\hat{F}^* = \hat{F}$. In the former statement,

$y = \hat{F} + Q^{-1} \bigtriangledown g(\hat{F})$, in which $\bigtriangledown g(F)$ is the derivative vector with respect to g at F (Groeneboom and Wellner, 1992).

The second condition concerns maximization of a quadratic function over region W. Suppose $\hat{F}^* = (\hat{F}_1^*, ..., \hat{F}_{r-1}^*)'$ has the same definition as before in W and define $Q = diag(q_j)$ a positive definite diagonal matrix. Furthermore, we let $P_0 = (0, 0)$ and

$$P_m = (\sum_{i=1}^{m} q_i, \sum_{i=1}^{m} q_i y_i), 1 \leq m \leq r-1,$$

be the points in $\Re^2$ for the constants $y = (y_1, ..., y_{r-1})' \in \Re^{r-1}$. The vector which contains the values of points $\{P_m; m = 0, ..., r-1\}$ is commonly referred to as a cumulative sum diagram because the coordinates of P_m is obtained by adding up the vectors $(q_1, ..., q_{r-1})'$ and $(q_1 y_1, ..., q_{r-1} y_{r-1})'$ cumulatively. Then $\hat{F}_i^*$ is considered as the left derivative of the convex minorant of, i.e. the largest convex function below the cumulative sum diagram $\{P_m; m = 0, ..., r-1\}$ evaluated at P_j.

While the first condition reveals the equivalence between the maximization of a general function $g(F)$ and the maximization of the quadratic function $g^*(F)$ for a given $\hat{F}$, the second condition pinpoints the location at which the maximization has occurred for a special quadratic function. Together, the two conditions motivate the ICM algorithm as follows:

Step 1. Select an initial estimate $\hat{F}_{C'}^0$ of $F_{C'}$.

Step 2. At the lth iteration, let the updated estimate $\hat{F}_{C'}^{(l)}$, which is equal to $(\hat{F}^{(l)}(C'_{(1)})$, ..., $\hat{F}^{(l)}(C'_{(r-1)})$, of $F_{C'}$ be the $\hat{F}^*$ that maximizes $g^*(F|y, Q(\hat{F}_{C'}^{(l-1)}))$ with

$$y = \hat{F}_{C'}^{(l-1)} - Q^{-1}(\hat{F}_{C'}^{(l-1)}) \bigtriangledown l_W(\hat{F}_{C'}^{(l-1)})$$

and $Q(\hat{F}_{C'}^{(l-1)})$ being positive definite diagonal matrix that may depend on $\hat{F}_{C'}^{(l-1)}$, where $l_W(F_{C'})$ is the log-likelihood function. In other words, $\hat{F}_{C'}^{(l)}$ is taken to be the derivative of the convex minorant of the cumulative sum diagram $\{P_m; m = 0, ..., r-1\}$ given by

$P_0 = (0,0)$ and

$$P_m = (\sum_{i=1}^{m} q_i^{(l-1)}, \sum_{i=1}^{m} (q_i^{(l-1)} \hat{F}^{(l-1)}(C'_{(i)}) - \frac{\partial}{\partial F(C'_{(i)})} l_W(\hat{F}_{C'}^{(l-1)})))$$

for $1 \leq m \leq r-1$, where $q_i^{(l-1)}$ is the ith diagonal entries of $Q(\hat{F}_{C'}^{(l-1)})$. Note that, in the ICM algorithm, a natural choice for $Q(F_{C'})$ is to take

$$q_j = q_j(F_{C'}) = -\frac{\partial^2}{\partial F^2(C'_{(j)})} l_W(F_{C'}),$$

assuming it exists for $j = 1, ..., r-1$.

Step 3. Return to step 2 until the algorithm converges.

Jongbloed (1998) shows that there exists problems associated with increments of log-likelihood and global convergence in the ICM algorithm, and suggests to add a line search into the algorithm based on the following fact.

Let g, g^*, W and $\hat{F}$ be as given in the first condition. For given F and a positive definite diagonal matrix $Q(F)$ that may depend on F, also let $A(F)$ be vector z at which $g^*(z|y, Q)$ achieves its maximum with

$$y = x - Q^{-1}(F) \bigtriangledown l_W(F).$$

Then for $F \neq \hat{F}$ and all arbitrary $\lambda > 0$,

$$g(F + \lambda(A(F) - F)) > g(F).$$

This fact shows once a line search is added into the algorithm, it will guarantee the log-likelihood function increases and the algorithm globally converges.

Define $0 < \epsilon < 0.5$ a constant that controls the line search process. The following step can be inserted between steps 2 and 3 of the ICM algorithm.

Step 2.1. If

$$l_W(\hat{F}_{C'}^{(l)}) > l_W(\hat{F}_{C'}^{(l-1)}) + (1-\epsilon)[\bigtriangledown l_W(\hat{F}_{C'}^{(l-1)})]'(\hat{F}_{C'}^{(l)} - \hat{F}_{C'}^{(l-1)}),$$

then the algorithm proceeds to step 3. Otherwise, find a point z such that

$$z = \hat{F}_{C'}^{(l-1)} + \lambda(\hat{F}_{C'}^{(l)} - \hat{F}_{C'}^{(l-1)})$$

for $0 \leq \lambda \leq 1$ that satisfies

$$\begin{aligned} \epsilon[\nabla l_W(\hat{F}_{C'}^{(l-1)})]'(z - \hat{F}_{C'}^{(l-1)}) &\leq l_W(z) - l_W(\hat{F}_{C'}^{(l-1)}) \\ &\leq (1-\epsilon)[\nabla l_W(\hat{F}_{C'}^{(l-1)}]'(z - \hat{F}_{C'}^{(l-1)}) \end{aligned}$$

Let $\hat{F}_{C'}$ denote the estimate given by the ICM algorithm. Then the NPMLE of F is given by $\hat{F}(t) = \hat{F}(C'_{(j)})$ if $C'_{(j)} \leq t < C'_{(j+1)}$ for $j = 0, ..., r-1$.

Aragon and Eberly (1992) discovered that one could have difficulty with the selection of initial values for the iterative convex minorant algorithm achieving global convergence. From the experiments they did, the results reveal that the steplength of the algorithm is too large in some precedents. Although Groeneboom (1990) considered using a buffer to prevent F or $1-F$ from being negative, this adjustment has a lack of available convergence results. Therefore, in order to resolve these issues, they developed a modified version of the algorithm and namely, a damped iterative convex minorant algorithm.

The damped iterative convex minorant algorithm is defined as the following

$$F_{C'}^{(k+1)} = F_{C'}^{(k)} - \alpha^k B(F_{C'}^{(k)})^{-1} l'(F_{C'}^{(k)}) \tag{3.3}$$

where $B(F_{C'}^{(k)}) = diag(-l''_{ii}(F_{C'}^{(k)}))$. When $F_{C'}^{(k)}$ approaches the maximal point $F_{C'}^{*}$ (at which the maximization of log-likelihood has achieved) very closely, the damped ICM algorithm will eventually become the regular ICM algorithm. To maintain the proper status of F as being a cumulative distribution function, the steplengths α^k are small initially to ensure that the range of it will be bounded within the interval, $[0, 1]$. A few methods will be discussed later on obtaining the steplength variable, which will satisfy $\alpha^k \to 1$ as $k \to \infty$.

For the exposition of global convergence, Aragon and Eberly (1992) used results from Ortega and Rheinholdt (1970, p.502) and provided the following theorem.

Theorem 2. (Aragon and Eberly, 1992) *Assume that $g : D \subset R^r \to R^r$ is continuously differentiable on the open set D, that there is an $F^0_{C'} \in D$ such that $L := L(g(F^0_{C'}))$ is compact, and that g has a unique critical point $F^*_{C'} \in L$. Suppose that $A(F_{C'}) \in R^{r\times r}$ is positive definite for all $F_{C'} \in L$. Then the iterates*

$$F^{(k+1)}_{C'} = F^{(k)}_{C'} - \alpha^k A(F^{(k)}_{C'})^{-1} g'(F^{(k)}_{C'}), k \geq 0$$

converge to $F^*_{C'}$ for any sequences of α_k such that $F^{(k)}_{C'} \subset L$ and $\lim_{k\to\infty} g'(F^{(k)}_{C'})^T p^{(k)}/\|p^{(k)}\|$ $= 0$ where $p^{(k)} = A(F^{(k)}_{C'})^{-1} g'(F^{(k)}_{C'})$.

Let $g(F_{C'}) = -l(F_{C'})$, $D = R^0_{\max}$ which contains all the maximal points of l and $A(F_{C'}) = \mathrm{diag}(-l''_{ii}(F_{C'}))$, respectively. Then, one could follow the mechanism given in (3.3) to run the algorithm. Because g is strictly convex, L in the statement of theorem is compact for any $F^{(0)}_{C'} \in R_{\max}$. A number of ways for computing steplengths are available to ensure that $F^{(k)}_{C'} \subset L$. Among them, the Goldstein-Armijo algorithm is probably the one that has been used mostly (Ortega and Rheinboldt 1970). In the application of damped ICM algorithm, Aragon and Eberly (1992) selected

$$\alpha^k = \max\{(1/2)^m : m \geq 0, F^{(k)}_{C'} - (1/2)^m B(F^{(k)}_{C'})^{-1} l'(F^{(k)}_{C'}) \in L(g(F^{(k)}_{C'}))^o\},$$

and set $F^{(k+1)}_{C'} = F^{(k)}_{C'} - \alpha_k B(F^{(k)}_{C'})^{-1} l'(F^{(k)}_{C'})$.

The damped ICM algorithm will not only achieve its global convergence by starting with any initial values of $F_{C'}$ within a close set, W; more importantly, when the initial value exceeds W, the convex minorant portion of the algorithm should also be able to project it onto the boundary of W in the later iterates. Accordingly, Aragon and Eberly (1992) developed an alternative iteration for the algorithm which resulted in the following

expression.

$$\zeta = F_{C'}^{(k)} - \alpha^k(\text{diag}(-l''(F_{C'}^{(k)})))^{-1}l'(F_{C'}^{(k)}), \quad F_{C'}^{(k+1)} = \zeta \qquad \text{if } \zeta \in T^o,$$
$$= \text{Proj}(\zeta) \qquad \text{if } \zeta \notin T^o.$$

Furthermore, it guarantees that $-l(F_{C'}^{(k+1)}) > -l(F_{C'}^{(k)})$ when the number of k increases.

Pan (1999) pointed out that the damped iterative convex minorant algorithm is in fact a generalized gradient projection (GGP) scheme. Moreover, it can be regarded as a Newton-type iterative algorithm in the problem of constrained optimization (Mangasarian 1996; Bertsekas 1982).

Based on his derivation, suppose we want to maximize the log-likelihood function l in a closed convex set X. The algorithm will then iterate by

$$F_{C'}^{(m+1)} = \text{Proj}[F_{C'}^{(m)} + \alpha^{(m)}H^{(m)^{-1}} \bigtriangledown l(F_{C'}^{(m)}), H^{(m)}, W] \tag{3.4}$$

(Pan, 1999), where $\bigtriangledown l$ represents the first derivative of l with respect to $F_{C'}$ and H is taken by the negative second order derivatives, which is often referred as a symmetric positive definite matrix for letting l be strictly concave. The projection operation, Proj, is defined as

$$\text{Proj}[y, H, W] := \arg\min_{F_{C'}}\{(y - F_{C'})'H(y - F_{C'}) : F_{C'} \in W\}.$$

Then, the desired solution of $F_{C'}$ can be obtained once the algorithm converges at the maximum value of l.

Bertsekas (1982) and discussed the theoretical properties of the GGP. Among them, three of the important ones are (Pan, 1999),

$I. F_{C'}^* \in \arg\max_{F_{C'} \in W}\{l(F_{C'})\}$;

$II. F_{C'}^* = \text{Proj}[F_{C'}^* + \alpha H^{-1} \bigtriangledown l(F_{C'}^*), H, W]$ for any $\alpha > 0$; and

$III. F_{C'}^* = \text{Proj}[F_{C'}^* + \alpha H^{-1} \bigtriangledown l(F_{C'}^*), H, W]$ for some $\alpha > 0$.

He argued that if l can be differentiated with respect to $F_{C'}$, then condition II can be fulfilled by condition I, whereas if l is convex, condition III can deduce condition I.

Also, if we neglect the second order information of $l(F^*_{C'})$ and reduce H to be an identity matrix, then the generalized gradient projection (GGP) method will be degenerated as the gradient projection (GP) method. For other theoretical discussions about GGP (e.g. its linear convergence and superlinear convergence rate), one could refer to Polak (1971).

Although Zhan and Wellner (1995) discussed some potential problems that might exist in the proof of Aragon and Eberly (1992), the damped iteratvie convex minorant algorithm revealed its superiority in global convergence and likelihood increment as opposed to the original iterative convex minorant algorithm. Naturally, it becomes one of the ideal options which can be applied for estimating the distribution function in the analysis of semiparametric model. In the following section, we will give a detailed discussion on how the damped iterative convex minorant algorithm is implemented in the estimation of the additive hazards regression model.

Chapter 4

Additive Hazards Regression Model for Current Status Data

In the previous sections, we discussed some of the applications from which current status data arise. These applications often include animal tumorigenicity experiment, demographical surveys and epidemiology studies. In situations where tumorigenicity experiment is conducted, current status data as composed of tumor onset time is the only available information about underlying survival variables of interest (Dinse and Lagakos, 1983). In other situations such as cross-sectional studies, current status data offer relatively simple but rather reliable information about occurrences of the event than that complete data provide. One example about such case is in epidemiological studies, we want to examine whether certain chronical diseases are presented within a specific time period (Keiding, 1991; Keiding et al., 1996; Shiboski and Jewell, 1992). Another example would be the pregnancy or marriage surveys in demographical studies.

For the past two decades, there has been a tremendous amount of research in the analysis of current status data. For instance, the articles that studied the proportional hazards model including Huang (1996), and Huang and Wellner (1997) where they developed a profile maximum likelihood approach to estimate the parameters of interests. Huang (1995) and Rossini and Tsiatis (1996) investigated the proportional odds model by using sieve maximum likelihood method, and the authors who discussed the additive hazards model for current status data include Ghosh (2001), Lin et al. (1998) and Martinussen and Scheike (2002b). Moreover, Sun (2005) studied the linear transformation model and developed some estimating equation approaches for estimation of regression

parameters. Shiboski (1998) proposed some generalized additive models and applied maximum likelihood approach with the use of step function approximation for inference. Other models that have been researched include the accelerated failure time model (Shen, 2000; Xue et al., 2004), the binary choice model (Huang and Wellner, 1996; Klein and Spady, 1993), generalized linear models (Jewell and Shiboski, 1990), and spline models (Grummer-Strawn, 1993).

The objective of this chapter is to discuss the estimation as well as inference procedures for some commonly used semi-parametric regression models, merely the additive hazards regression model, involved in the analysis of current status data. In this regard, one of the well-known approaches is the maximum likelihood estimation. Although this is rather a quite common method, the likelihood in the current status data yet contains both finite-dimensional regression parameters and infinite-dimensional nuisance parameters (e.g. the cumulative baseline hazard or cumulative distribution function). It is not easy to achieve by using the approach. Apparently, one of the means that yields the solution for such case is to estimate the regression parameters and nuisance terms simultaneously by the damped ICM algorithm introduced in Chapter 3. This approach differs from the method that we often used to estimate the proportional hazards model for right-censored survival data because the latter method only deals with partial likelihood that does not involve any of the nuisance parameters and the properties associated with the parameters of interest can be easily derived by martingale theory. However, in current status data, the partial likelihood fails to capture the function of the nuisance parameters and therefore, is not suitable for the specified estimation procedure. Instead, one has to work with the full likelihood. In the following sections, we will describe these procedures for specific semi-parametric models in detail.

4.0.7 Proportional Hazards Regression Model

This section will introduce analysis principles of using the proportional hazards model in current status data. This model was first proposed by Cox (1972) and was developed in order to estimate the effects of different covariates influencing the time-to-event of a system. The model has been widely used in the biomedical field and recently, there has been an increasing interest in its application in reliability engineering. The basic form of the model is specified as follows:

$$\lambda(t|Z) = \lambda_0(t)\exp(\beta^t Z), \tag{4.1}$$

where $\lambda_0(t)$ is an arbitrary baseline hazard rate function of t, Z is a vector of covariates and $\beta = (\beta_1, ..., \beta_p)^t$ is a parameter vector.

The likelihood function is proportional to

$$L(\beta, \Lambda_0) = \prod_{i=1}^{n} \{[1 - \exp(-\Lambda_0(C_i)e^{\beta^t Z_i})]^{\delta_i} \cdot [\exp(-\Lambda_0(C_i)e^{\beta^t Z_i})]^{1-\delta_i}\} \tag{4.2}$$

where Λ_0 is the cumulative baseline hazard function, which is defined as $\Lambda_0(t) = \int_0^t \lambda_0(s)\, ds$ and $\delta_i = I\{C_i \geq T_i\}$.

In terms of β and F_0, the baseline distribution function, the likelihood can be rewritten as

$$L(\beta, F_0) = \prod_{i=1}^{n} \{[1 - (1 - F_0(C_i))^{e^{\beta^t Z_i}}]^{\delta_i} \cdot [(1 - F_0(C_i))^{e^{\beta^t Z_i}}]^{1-\delta_i}\}. \tag{4.3}$$

Maximum likelihood estimation are used to estimate β and F_0 by calculating the maximum value of the likelihood function $L(\beta, F_0)$ given in (4.3). For this and a given set of current status data, only the values of $F_0(t)$ at the monitoring time C_i's affect the likelihood function. Hence, without loss of generality, one can focus only on the maximization of $L(\beta, F_0)$ over all nondecreasing stepwise functions with jumps only at the C_i's for $F_0(t)$.

Let $0 < C_1 < C_2 ... < C_m$ be the ordered distinct time point of $\{C_i\}_{i=1}^n$ and Ω_F, the set of all baseline cumulative distribution functions $F_0(t)$. Thus the log-likelihood of Cox proportional hazards model for current status data has the form

$$l(\beta, F_0) = \sum_{i=1}^{n} \{\delta_i \log[1 - (1 - F_0(C_i))^{e^{\beta^t Z_i}}] + (1 - \delta_i)\log[(1 - F_0(C_i))^{e^{\beta^t Z_i}}]\}. \quad (4.4)$$

Then, the maximization of $l(\beta, F_0)$ can be solved by using the Newton-Raphson algorithm which requires the corresponding partial derivatives of the log-likelihood function with respect to the parameter space β and the cumulative distribution function F_0, respectively.

4.0.8 Additive Hazards Regression Model

As with the proportional hazards model, the additive hazards model specifies the effects of covariates on the failure time through the hazard function. Particularly, it assumes that the hazard function of T at time t, given the history of a p-dimensional covariate process $Z(.)$ up to t, has the form

$$\lambda(t|Z) = \lambda_0(t) + \beta^t Z(t), \quad (4.5)$$

where $\lambda_0(t)$ is an unspecified baseline hazard function, and β stands for a p-vector of unknown regression parameters of interests (Lin et al., 1998). Overall, The effects of covariates in such regression model are to additively increase or decrease the hazard function.

Although both the proportional hazards model and the additive hazards model concentrate their forms on hazard function, the effects of covariates have different meanings. Under the proportional hazards model, the parameters of interests β denote the logarithm of the ratio between risk factors and failure rates, whereas in the additive hazard models, β represents the difference of that risk with respect to such factors and failure

rates. This can be easily seen at the situation where Z takes only binary values, 0 or 1. In this case, we have

$$\lambda(t|Z=1) = \lambda(t|Z=0) + \beta.$$

One attractive feature of the additive hazards model is that it provides a relatively simple method for analyzing survival data when the latent variables or frailties are present. For the additive frailty model, the marginal model is still in the form of the additive hazards model and the parameter of interests β has the identical meanings in both the additive frailty model and the marginal model (Lin, Oakes and Ying, 1998; Lin and Ying, 1997). However, the result will not be true when the similar setting applies to the proportional hazards model.

There exists an extensive literature on the theoretical discussions and applications of the additive hazards model. Sources include Breslow and Day (1987), Kim and Lee (1998), Kulich and Lin (2000), and Lin and Ying (1994). Some methodologies of the additive hazards model have been developed to make it more flexible when applying it to the right-censored survival data. For instance, we could let Z be time-dependent, and as a result, inferences about β can be derived similarly to those when Z is time-independent. Lin and Ying (1995) introduced an additive-multiplicative hazard model which combines the proportional hazards model and the additive hazards model together. Martinussen and Scheike (2002a) and Scheike and Zhang (2002) provide developments of the models in a further depth.

As discussed previously, the additive hazards model (4.5) is another commonly used semi-parametric regression model in survival analysis in addition to the Cox proportional hazards model. For the purpose of fitting and making inference on it to the current status data, we assume that given a set of covariates Z, the model can be expressed, in terms

of the cumulative distribution function, as

$$F(t|Z) = 1 - [1 - F_0(t|Z)]\exp(-\beta^t Z(t)). \tag{4.6}$$

Let $0 < C_1 < C_2 ... < C_m$ be the ordered distinct time points of $\{C_i\}_{i=1}^n$ and Ω_F the set of all baseline cumulative distribution functions $F_0(t)$. Thus, the log-likelihood function of the additive hazards model for current status data is proportional to

$$l(F_0, \beta) = \sum_{i=1}^{n}(1-\delta_i)\log((1-F_0(C_i))e^{-\beta^t Z_i^*(C_i)}) + \sum_{i=1}^{n}\delta_i \log(1-(1-F_0(C_i))e^{-\beta^t Z_i^*(C_i)}), \tag{4.7}$$

where $\delta_i = I\{C_i \geq T_i\}$ and $Z_i^*(C_i) = Z_i \cdot C_i$. The main advantage of this parametrization is that the log-likelihood function is concave with respect to the cumulative baseline distribution function, F_0.

To estimate F_0 and β, a natural approach is to use the nonparametric maximum likelihood estimation. For its implementation, we adopt the method developed by Aragon and Eberly (1992) and Pan (1999), which yields a process that is to apply the damped iterative convex minorant (ICM) algorithm in the formulation of generalized gradient projection (GGP) to generate the estimates.

So far for what we have discussed, the damped ICM algorithm (Aragon and Eberly, 1992) is only available for current status data without covariates. The original version of the ICM algorithm developed by Groeneboom and Wellner (1992) was illustrated in the form of stochastic processes. We transform it into a GGP scheme (Pan 1999; Mangasarian 1996 and Bertsekas 1982) and extend it to the situation where both cumulative distribution function and covariates exist so that it can be used to estimate these parameters simultaneously. By doing this, its connection with the Newton-Raphson algorithm becomes quite obvious because, if we focus our attentions on the inner part of the iteration which is included by the projection operation, that expression is exactly the Newton-Raphson iteration. Estimating parameters of covariates, β, only requires the

use of the Newton-Raphson iteration; since there is no constraints on β, the projection operation becomes unnecessary.

In general, the damped ICM algorithm needs the first and second derivatives of the log-likelihood function to estimate the parameters of interest. Now let ∇_1 and ∇_2 be the first derivatives with respect to F_0 and β, respectively. Then,

$$\begin{aligned}\nabla_1 l(F_0,\beta) &= \partial l(F_0,\beta)/\partial F_0 \\ &= \sum_{i=1}^{n}\{(1-\delta_i)\frac{-1}{1-F_0(C_i)} + \delta_i \frac{e^{-\beta^t Z_i^*(C_i)}}{1-(1-F_0(C_i))e^{-\beta^t Z_i^*(C_i)}}\}\end{aligned}$$

and

$$\begin{aligned}\nabla_2 l(F_0,\beta) &= \partial l(F_0,\beta)/\partial \beta \\ &= \sum_{i=1}^{n}\{(1-\delta_i)(-Z_i^*(C_i)) + \delta_i \frac{(1-F_0(C_i))e^{-\beta^t Z_i^*(C_i)}(Z_i^*(C_i))}{1-(1-F_0(C_i))e^{-\beta^t Z_i^*(C_i)}}\}.\end{aligned}$$

Similarly, denote $G_1(F_0,\theta)$ and $G_2(F_0,\theta)$ to be the corresponding diagonal matrices of the negative seccond order derivatives. Therefore,

$$\begin{aligned}G_1(F_0,\beta) &= -\partial^2 l(F_0,\beta)/\partial F_0^2 \\ &= \sum_{i=1}^{n}\{(1-\delta_i)\frac{1}{(1-F_0(C_i))^2} + \delta_i \frac{e^{-2\beta^t Z_i^*(C_i)}}{[1-(1-F_0(C_i))e^{-\beta^t Z_i^*(C_i)}]^2}\}\end{aligned}$$

and

$$\begin{aligned}G_2(F_0,\beta) &= -\partial^2 l(F_0,\beta)/\partial \beta^2 \\ &= \sum_{i=1}^{n}\{0 + \delta_i[\frac{[(1-F_0(C_i))e^{-\beta^t Z_i^*(C_i)}Z_i^*(C_i)]^2}{[1-(1-F_0(C_i))e^{-\beta^t Z_i^*(C_i)}]^2} \\ &\quad + \frac{[(1-F_0(C_i))e^{-\beta^t Z_i^*(C_i)}Z_i^{*2}(C_i)]}{[1-(1-F_0(C_i))e^{-\beta^t Z_i^*(C_i)}]}]\}.\end{aligned}$$

Then the damped ICM algorithm will iterate as

$$\begin{aligned}F_0^{(k+1)} &= \mathrm{Proj}[F_0^{(k)} + \alpha^{(k)} G_1(F_0^{(k)},\beta^{(k)})^{-1}\nabla_1 l(F_0^{(k)},\beta^{(k)}), G_1(F_0^{(k)},\beta^{(k)}), R], \\ \beta^{(k+1)} &= \beta^{(k)} + \alpha^{(k)} G_2(F_0^{(k)},\beta^{(k)})^{-1}\nabla_2 l(F_0^{(k)},\beta^{(k)}).\end{aligned}$$

where

$$\alpha^{(k)} = \max\{1/2^i : l(F_0^{(k+1)}, \beta^{(k+1)}) > l(F_0^{(k)}, \beta^{(k)}), i = 0, 1, 2, ...\}$$

and Proj is defined as in Chapter 3 to ensure that the estimate of F_0 is a proper distribution function (i.e. nondecreasing and ranging between 0 and 1).

In addition to the above procedures, one also needs to select initial values of β and F_0 in order to start the algorithm. A natural choice for obtaining F_0 is to use a stepwise distribution function. However, it is optional that one could also adopt the derivations developed by Lin and Ying (1998) which will be introduced later in this section.

As Pan (1999) pointed out, despite the trivial process for inverting G_1 and G_2, to avoid the occurrence of zeros in the diagonal entries, the Levenberg-Marquardt adjustment is used (Thisted 1988). Specifically, G_1^{-1} and G_2^{-1} are replaced by $(G_1 + \epsilon_0 I)^{-1}$ and $(G_2 + \epsilon_0 I)^{-1}$ with a carefully selected arbitrary number $\epsilon > 0$ when they are near singular. By intuition, the damped ICM algorithm will converge, as the log-likelihood function has an upper limit and each step of the modified algorithm increases the log-likelihood. If we ignore the second order information and let G be an identity matrix, the GGP will simply become the gradient projection (GP), which only yields a linear converge rate. But, if G is carefully selected (e.g. to be partially diagonal), then the GGP can have a superlinear convergence rate (Bersekas 1982. Propositions 3 and 4). Since the Hessian matrix G in our ICM algorithm is between them, we can expect the convergence rate of the algorithm will lie between linear and superlinear; moreover, if the diagonal elements of the Hessian matrix exceed the nondiagonals, the convergence rate of the algorithm will be minimally close to superlinearity.

We remark that although the damped ICM algorithm generally works well, the computation involved could become intensive, and one may face unstable estimation problems for some data sets such as those that have a large number of different observation time points. As an alternative, for a given data set, one could maximize the log-likelihood

function $l(\beta, \Lambda_0)$ over β and Λ_0 in lieu of $l(\beta, F_0)$. As a result, the log-likelihood function takes the form

$$l(\Lambda_0, \beta) = \sum_{i=1}^{n} \{(1-\delta_i)\log(e^{-\Lambda_0(C_i)-\beta^t Z_i^*(C_i)}) + \delta_i \log(1 - e^{-\Lambda_0(C_i)-\beta^t Z_i^*(C_i)})\}.$$

It can be shown that $l(\Lambda_0, \beta)$ is concave with respect to Λ_0 for given β. Huang (1996) and Huang and Wellner (1997) studied a similar situation for the Cox proportional hazards model and suggested a profile maximum likelihood approach to estimate parameters of interests. It is likely that their estimation method can be also applied to the additive hazards regression model.

Lin, Oakes and Ying (1998) developed a simple equation approach to estimate the additive hazards regression model. To have a better understanding on their method, let us assume that for any given set of current status data, the monitoring time C is independent of the failure time T and covariates Z. Then, following the discussions we had in Chapter 2, the counting process for the number of observations is defined by $N_i(t) = I[C_i \leq \min(T_i, t)]$, which jumps by one whenever the subject i where $i = 1, ..., N$ is surveilled at time t and found to be failure free. $N_i(t)$ is considered as censored if a subject i has experienced failure under the surveillance at time t. It follows that the intensity function $dN_i(t)$ takes the form

$$dH_i(t; Z_i) = Y_i(t)\exp(-\beta^t(Z_i^*(t)))dH_0(t), \tag{4.8}$$

where $Y_i(t) = I(C_i \geq t), Z_i^*(t) = \int_0^t Z_i(t)dt, dH_0(t) = \exp(-\Lambda_0(t))d\Lambda_c(t)$ with $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ and $\Lambda_c(t) = \int_0^t \lambda_c(s)ds$. Thus, equation (4.8) yields the form of the Cox proportional hazards model. In the meantime, it suggests that the counting process

$$M_i(t) = N_i(t) - \int_0^t Y_i(s)\exp(-\beta^t(Z_i^*(t))dH_0(s) \tag{4.9}$$

is a martingale defined in the σ-filtration

$$\mathcal{F}_t = \sigma\{N_i(s), Y_i(s), Z_i^*(s) : s \leq t, i = 1, ..., n\}.$$

Hence, one can apply partial likelihood approach to model (4.8) for inference about β.

Suppose

$$S^{\otimes m}(t;\beta) = \sum_{i=1}^{N} Y_i(t)\exp(-\beta^t(Z_i^*(t)))(Z_i^*(t))^{\otimes m},$$

where $Z_i^*(t)^{\otimes 0} = 1, Z_i^*(t)^{\otimes 1} = Z_i^*(t)$ and $Z_i^*(t)^{\otimes 2} = Z_i^*(t)Z_i^*(t)'$ up to $m = 0, 1, 2$. By the partial likelihood approach, the score function is yielded as

$$U(\beta) = \sum_{i=1}^{N} \int_0^{\infty} \left[Z_i^*(t) - \frac{S^{(1)}(t;\beta)}{S^{(0)}(t;\beta)} \right] dN_i(t). \quad .$$

Similarly, the information matrix for β has the form

$$I(\beta) = \sum_{i=1}^{N} \int_0^{\infty} \left[\frac{S^{(2)}(t;\beta)}{S^{(0)}(t;\beta)} - \frac{S^{(1)}(t;\beta)^{\otimes 2}}{S^{(0)}(t;\beta)^2} \right] dN_i(t).$$

Then one can estimate β by $\hat{\beta}$ defined as the solution to $U(\beta) = 0$ by the Newton-Raphson algorithm. It can be shown that under some regularity conditions, $\hat{\beta}$ is consistent (Kalbfleisch and Prentice, 2002; Lin, Oakes and Ying, 1998). Also for large n, the distribution of $\sqrt{n}(\hat{\beta} - \beta)$ and $\sqrt{n}U(\beta)$ can be approximated by the multivariate normal distribution with mean zero and variance-covariance matrices Ω^{-1} and Ω, respectively, where $\Omega = \lim n^{-1} I(\beta)$.

It is required that one needs to input the initial vector of β and Λ, namely, β_0 and $\Lambda_0(t)$, to utilize the Newton-Raphson algorithm. A natural option for β_0 is to use the results directly produced from the Cox PH model, whereas for Λ_0, Lin, Oakes and Ying (1998) proposed that it could be obtained by solving the equation $\hat{H}(t) = \int_0^t \exp(-\Lambda_0(u))d\hat{\Lambda}_{c,0}(u)$, where $\hat{H}$ and $\hat{\Lambda}_{c,0}$ are the Aalen-Breslow type estimators of $H(t)$ and $\Lambda_{c,0}$ given by

$$\hat{H}(t) = \sum_{i=1}^{n} \int_0^t \frac{dN_i(s)}{\sum_{j=1}^{n} Y_j(s)\exp(-\beta_0 Z_j^*(s))}$$

and

$$\hat{\Lambda}_{c,0}(t) = \sum_{i=1}^{n} \int_0^t \frac{dN_i^c(s)}{\sum_{j=1}^{n} Y_j(s)}.$$

Here, $N_i^c(s) = I(C_i \leq s)$ and it is the complement of the counting process $N_i(s)$ defined previously. Note that Lin et al.(1998)'s proposition on estimating initial guess of $\Lambda_0(t)$ can be also used in our damped ICM algorithm.

It can be seen that Lin et al. (1998)'s approach here eventually transforms the analysis problem to regression analysis of right-censored survival data using the proportional hazards model, which can be quite easily performed. As a result, one can apply the pre-existing software for the proportional hazards model to estimate the parameters of the additive hazards model.

Note that unlike the approaches we have given in the previous section, the above estimation procedure requires that the C_is follow the proportional hazards model. This may be restrictive and could lead in biased estimates of regression parameters if the model is incorrect. In general, an estimating equation approach has the disadvantage that it may not be as efficient as the damped ICM algorithm approach. This is true because the distribution of the censoring time for $N_i(t)$ contains regression parameter β (Lin, Oakes, and Ying, 1998; Martinussen and Scheike, 2002b), which are usually informative and the method discussed here represents a trade-off between simplicity and efficiency and the amount of efficiency loss under specific circumstances.

In addition to our damped ICM and Lin et al. (1998)'s approaches, one can also apply other methods to estimate β for additive hazards regression. For example, Martinussen and Scheike (2002b) developed an efficient score function for β, which has the form

$$U_E(\beta;\Lambda_0) = \sum_{i=1}^{n} \int_0^{\infty} \left[Z_i^*(t) - \frac{S_E^{(1)}(t;\beta,\Lambda_0)}{S_E^{\otimes 0}(t;\beta,\Lambda_0)} \right] \times \left\{ \frac{\exp[-\Lambda_0(t) - \beta^t(Z_i^*(t))]}{1-\exp[-\Lambda_0(t) - \beta^t(Z_i^*(t))]} dN_i^*(t) - dN_i(t) \right\}$$

where $N_i^*(t) = I(C_i \leq t) - N_i(t)$ and

$$S_E^{\otimes m}(t;\beta,\Lambda_0) = \sum_{i=1}^{N} Y_i(t)\alpha_i^c(t;Z_i^*) \frac{\exp(-\Lambda_0(t) - \beta^t(Z_i^*(t)))}{1-\exp(-\Lambda_0(t) - \beta^t(Z_i^*(t)))} (Z_i^*(t))^{\otimes m}.$$

In the above expression, $\alpha_i^c(t; Z_i^*)$ represents the hazard function of C_i given Z_i^*, which does not have to follow model (4.8). To apply this approach, one needs to estimate both $\Lambda_0(t)$ and $\alpha_i^c(t; Z_i^*)$ initially, then β can be obtained from the solution $U_E(\beta; \Lambda_0) = 0$ where both $\Lambda_0(t)$ and $U_E(\beta; \Lambda_0)$ are fulfilled by their estimates. Martinussen and Scheike (2002b) showed that the efficient estimator has the consistency property and follows a multivariate normal distribution asymptotically with covariance matrix reaching the information lower bound. However, this approach is very difficult to be materialized in computational program in comparison with our damped ICM algorithm.

Chapter 5

Numerical Studies

5.0.9 Simulation Results

In this section, we report the results of simulation studies. We used Monte Carlo methods to illustrate that the damped ICM algorithm provides good accuracy and efficiency for the estimation of the additive hazards model.

To evaluate our estimation procedure, we conducted the same simulation as Lin, Oakes and Ying proposed in their 1998's paper. According to their proposal, the failure times were produced based on the form of the model (4.5), in which the baseline hazard function λ_0 and the true value of parameter of covariate, β_0, were assigned to be 1 and 0.5, respectively; and the covariate, Z, is treated as a random variable following a uniform distribution between 0 and $\sqrt{12}$. Consider an exponential distribution with a scale parameter λ_c, thus one could obtain the examining time C_i by generating random numbers from such distribution with the scale parameter specified. In Lin, Oakes and Ying (1998), C was generated in accordance with $\lambda_c = 0.5$, 1.0 and 1.5, respectively. Sample sizes of 100 and 200 were considered and each group of the simulation parameters had run in 1,000 simulated samples.

All the programs, including damped ICM algorithm and simulation codes, have been implemented in R. The initial guess for the baseline cumulative distribution function is in a form of a stepwise discrete function which satisfies the conditions of monotone nondecreasing and being restricted in a range from 0 to 1. The convergence criterion, ϵ, was set to be equal to 10^{-3}. Therefore, the algorithm will be stopped once both the log-likelihood increment and the change of the regression coefficient from two consecutive iterations are less than that value.

During the process of simulation, we discovered that the damped ICM algorithm had a relatively faster convergence speed than that other algorithms achieved. This fact motivates us to attempt using some intensive computational methods, such as bootstrap, incorporated with the extended ICM algorithm to obtain the confidence intervals. The bootstrap is used to measure the variability of the NPMLE $\hat{\beta}$. For related work on this subject, one could refer to Efron and Tibshirani (1986).

Our bootstrap scheme is similar to what has been discussed in Burr (1994), where a bootstrap sample is formed by taking the same number of observations randomly with replacement from the original sample. We used 100 bootstrap samples and the results are reported in Table 5.1. Noticeably the NPMLE tends to be more biased when $\lambda_0 = 0.5$. The reason behind this might be that there is a tremendous amount of information had been lost from censoring, particularly a large proportion (almost 70 percent) left-censoring as opposed to roughly the rest 30 percent right-censoring within an individual data set. For a left censored observation $(0, C_i, Z_i)$, its contribution to the log-likelihood can be written as

$$L_i = \log\{1 - (1 - F(C_i))\exp(-\beta^t Z_i)\}$$

so that the maximization of it can be achieved when β approaches to the infinity with a nonnegative Z_i. On the other hand, the bootstrap estimate for the variance of the NPMLE is quite sensible. One could see the result based on the situation where $\lambda_0 = 1.5$, the proportion of left-censoring is reduced to 56 percent. Overall, the bootstrap estimate improves along with the NPMLE and we can thus claim that it works reasonably well.

As reviewed in Lin, Oakes and Ying (1998), their estimation method is to use the partial maximum likelihood approach applying for a right-censoring setting by alternating the form of the additive hazards regression into ordinary Cox proportional hazards model. In contrast, we propose to estimate the baseline distribution function and regression

	n=100			n=200		
	$\lambda_c = 0.5$	1.0	1.5	$\lambda_c = 0.5$	1.0	1.5
Mean of $\hat{\beta}$	0.51	0.51	0.50	0.51	0.50	0.50
Stand. error of $\hat{\beta}$	0.17	0.22	0.13	0.10	0.10	0.15
Bootstrap mean of SE($\hat{\beta}$)	0.19	0.25	0.17	0.12	0.14	0.18
Bootstrap S.D. of SE($\hat{\beta}$)	0.15	0.23	0.13	0.09	0.10	0.12
Cov. prob. of 95% CI	0.96	0.96	0.96	0.97	0.96	0.96

Table 5.1: NPMLE of the regression parameter, $\hat{\beta}$, and the bootstrap estimate of its standard error.

coefficients simultaneously from maximizing the full likelihood function by using the damped iterative convex minorant algorithm. We compare these two methods by testing each one of them in the same simulation environment and the results are shown in Table 5.2.

(a) Simulation results from the damped ICM algorithm

	n=100			n=200		
	$\lambda_c = 0.5$	1.0	1.5	$\lambda_c = 0.5$	1.0	1.5
Mean of $\hat{\beta}$	0.51	0.51	0.50	0.51	0.50	0.50
Stand. error of $\hat{\beta}$	0.17	0.22	0.13	0.10	0.10	0.15
Bootstrap mean of SE($\hat{\beta}$)	0.19	0.25	0.17	0.12	0.14	0.18
Bootstrap S.D. of SE($\hat{\beta}$)	0.15	0.23	0.13	0.09	0.10	0.12
Cov. prob. of 95% CI	0.96	0.96	0.96	0.97	0.96	0.96

(b) Simulation results from Lin et al.(1998)

	n=100			n=200		
	$\lambda_c = 0.5$	1.0	1.5	$\lambda_c = 0.5$	1.0	1.5
Mean of $\hat{\beta}$	0.56	0.54	0.53	0.53	0.52	0.51
Stand. error of $\hat{\beta}$	0.45	0.40	0.40	0.29	0.26	0.27
Mean of SE($\hat{\beta}$)	0.42	0.38	0.39	0.28	0.26	0.27
Cov. prob. of 95% CI	0.96	0.96	0.95	0.96	0.96	0.95

Table 5.2: Summary statistics for the comparisons of two methods between the damped ICM algorithm and Lin et al.(1998)

The comparison suggests that the bias of $\hat{\beta}$ estimated from the damped ICM algorithm is smaller than that obtained from Lin et al. (1998). The standard error estimate from

the damped ICM algorithm tends to be smaller as opposed to the one from Lin et al. (1998). Both methods are able to provide coverage probabilities close to the nominal level. As expected, the proposed algorithm provides better accuracy and efficiency than Lin, Oakes and Ying (1998) did.

In addition to the above results, more simulation studies are conducted. Here, the covariate Z is considered as a binary random variable generated from a binomial distribution with the probabilities equal to 0.5, 0.4 and 0.6, respectively. Keeping any other factors unchanged, we want to investigate how close the estimate is to the true parameter $\beta = 0.5$ and $\beta = 1$. The outputs are shown in the following tables.

The simulation yields similar results as those were discussed previously. The estimated standard errors tend to be higher in the case of sample size 100 but it is becoming lower in the case of sample 200. This might be due to the fact that less number of observations in the case of sample size 100 contributes information to the log-likelihood function as opposed to the case of sample size 200. When λ_c increases, the estimate, $\hat{\beta}$, tends to be closer to the true value, $\beta = 0.5$ and the corresponding estimated standard errors become smaller. This is because more amount of information has gained from right-censoring. Once we changed the probability, that result becomes clearer. From $p = 0.4$ to $p = 0.6$, the number of left-censored observations decrease dramatically, which, on the contrary, increases the number of right-censored observations and let the data set be able to contribute more information to the log-likelihood function. This provides a better condition for the damped ICM algorithm to perform its estimation power and helps the estimate of parameter eventually to become unbiased.

5.0.10 A Real Example

To illustrate the application of the damped ICM algorithm to Additive Hazards Regression model, we apply it into a real example. This example concerns the lung tumor data described in Chapter 1.

As mentioned before, lung tumors are usually considered as nonlethal, thus one can reasonably assume that the data has a current status characteristic with respect to the time to tumor onset. Also the death times, which are sometimes called observation times, can be assumed to be independent of tumor onset times within each treatment group. To compare the tumor incidence rates between the two treatments groups, we first note that for the data given in Table 1.2, the number of animals who had developed tumors at their deaths are 27 and 35 in the conventional and germ-free environments, respectively. This gives empirical tumor development rates of 0.28 and 0.73 without considering death time information and suggests that there is a difference between the tumor rates in the two groups.

Define $Z_i = 0$ for the animals in the conventional environment (CE) and 1 for those in the germ-free environment (GE). Also let T_i's be the occurrence times of lung tumors for the animals in the study and suppose that they can be described by the additive hazards regression model (4.5). For estimation of the effect of the environmental factor on tumor growth, the damped ICM algorithm approach gives $\hat{\beta} = 0.00073$ with estimated standard deviation equal to 0.00032357. This gives a p-value of 0.024 for comparison of the two groups, which indicates that the distributions of the observation times differ. Overall, the results here suggest that the animals in the germ-free environment had significantly higher lung tumor incidence than those in the conventional environment. Figure 5.1 also presents the NPMLE of the survival functions of times to lung tumor for animals in the two environmental groups based on the estimation procedure of the damped ICM algorithm.

The above data were also analyzed by Lin et al. (1998). They obtained a coefficient estimate of 0.00071 with an estimated standard error of 0.00041. Their z-test statistic was equal to 1.73 for testing the environmental difference in tumor incidence, which resulted a P-value of 0.084. Comparably, our method yields a stronger evidence for the difference between the two environmental impacts on tumor growth. This discrepancy of the results reveals that our damped ICM algorithm gained more efficiency than the approach of Lin et al did. (1998).

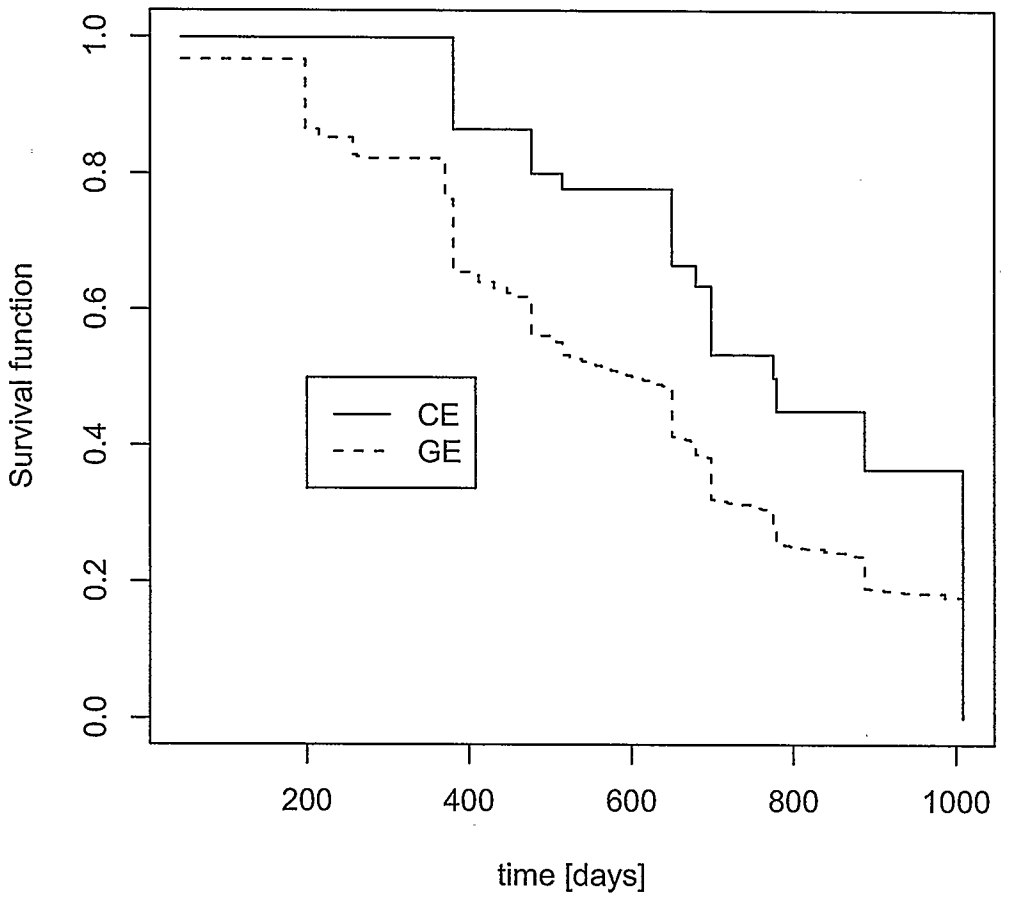

Figure 5.1: Estimates of survival functions of time to lung tumor onset for both conventional environment (CE) and germ-free environment (GE)

	n=100			n=200		
	$\lambda_c = 0.5$	1.0	1.5	$\lambda_c = 0.5$	1.0	1.5
Mean of $\hat{\beta}$	0.50	0.50	0.51	0.50	0.50	0.51
Stand. error of $\hat{\beta}$	0.26	0.27	0.30	0.23	0.26	0.26
Bootstrap mean of SE($\hat{\beta}$)	0.29	0.31	0.32	0.25	0.28	0.28
Bootstrap S.D. of SE($\hat{\beta}$)	0.24	0.27	0.28	0.22	0.25	0.25
Cov. prob. of 95% CI	0.95	0.95	0.95	0.96	0.95	0.95

Table 5.3: Simulation results for $\hat{\beta}$ when $p = 0.5, \lambda_0 = 0.5$ and $\beta = 0.5$.

	n=100			n=200		
	$\lambda_c = 0.5$	1.0	1.5	$\lambda_c = 0.5$	1.0	1.5
Mean of $\hat{\beta}$	0.51	0.51	0.51	0.50	0.50	0.51
Stand. error of $\hat{\beta}$	0.43	0.47	0.49	0.36	0.37	0.40
Bootstrap mean of SE($\hat{\beta}$)	0.45	0.49	0.50	0.38	0.38	0.43
Bootstrap S.D. of SE($\hat{\beta}$)	0.38	0.46	0.48	0.32	0.32	0.34
Cov. prob. of 95% CI	0.95	0.95	0.95	0.96	0.96	0.95

Table 5.4: Simulation results for $\hat{\beta}$ when $p = 0.4, \lambda_0 = 0.5$ and $\beta = 0.5$.

	n=100			n=200		
	$\lambda_c = 0.5$	1.0	1.5	$\lambda_c = 0.5$	1.0	1.5
Mean of $\hat{\beta}$	0.50	0.50	0.51	0.50	0.50	0.50
Stand. error of $\hat{\beta}$	0.24	0.25	0.27	0.21	0.21	0.22
Bootstrap mean of SE($\hat{\beta}$)	0.26	0.28	0.30	0.23	0.25	0.24
Bootstrap S.D. of SE($\hat{\beta}$)	0.23	0.23	0.23	0.20	0.19	0.20
Cov. prob. of 95% CI	0.96	0.95	0.95	0.96	0.96	0.96

Table 5.5: Simulation results for $\hat{\beta}$ when $p = 0.6, \lambda_0 = 0.5$ and $\beta = 0.5$.

	n=100			n=200		
	$p = 0.5$	0.4	0.6	$p = 0.5$	0.4	0.6
Mean of $\hat{\beta}$	0.99	0.98	0.99	1.00	0.99	1.00
Stand. error of $\hat{\beta}$	0.28	0.34	0.28	0.23	0.29	0.24
Bootstrap mean of SE($\hat{\beta}$)	0.32	0.39	0.37	0.23	0.26	0.24
Bootstrap S.D. of SE($\hat{\beta}$)	0.28	0.36	0.26	0.22	0.25	0.22
Cov. prob. of 95% CI	0.96	0.95	0.95	0.96	0.95	0.96

Table 5.6: Simulation results for $\hat{\beta}$ when $\lambda_c = 0.5, \lambda_0 = 0.5$ and $\beta = 1$.

Chapter 6

Conclusion

As what we have mentioned previously, there exists extensive literature about current status data in the context of demographical studies (Diamond and McDonald, 1991; Diamond et al., 1986) and tumorigenicity experiments (Dinse and Lagakos, 1983; Dewanji and Kalbleisch, 1986). Nevertheless, the literature about current status data from the areas of survival studies is limited, especially about modeling current status data under the commonly used semiparametric survival models. From the early discussion, the articles that highlighted rigorous studies for the use of the additive hazards regression model including Ghosh (2001), Lin et al. (1998) and Martinussen et al. (2002b). Among them, the asymptotic study of some of the other similar inference procedures are introduced. The authors who discussed the Cox proportional hazards model for current status data include Huang (1996) and Huang and Wellner (1997).

The purpose of this thesis is to demonstrate the estimation and inference procedures for additive hazards regression model in the context of current status data. By far, we have given a throughful discussion on types of methods for analyzing such model. Additionally, some of the essential concepts in survival analysis such as structures of survival data, basic quantities and counting processes as well as the illustration of Cox proportional hazards model, NPMLE and likelihood construction are also included. Through the comparisons in both theory and application (i.e. simulation and an example demonstration), the damped ICM approach, which has the least constraints and high efficiency, proved its most superiority among the other pre-existing methods. In the mean time, the fast convergence speed of the algorithm facilitates its use combing with other computing-intensive methods like bootstrap, so that one could obtain the empirical con-

fidence intervals and other related results easily and conveniently. Finally, its robustness on application reveals that the damped ICM algorithm can be easily extended to other types of interval-censored survival data.

Despite of many advantages by using the damped ICM algorithm, we are certain that it has some weaknesses as well. Particularly, an issue related to deal with high-dimensional parameter space. Although, in theory, the use of the complex Hessian matrix makes it possible to generate multiple estimates for the underlying model which contains more than one coefficients, by the time when we were running simulation, the Hessian matrix was difficult to maintain its positive definiteness and consequently, the algorithm was often unable to converge or produced the results which were far from what we were expecting. Therefore, it is necessary to have further investigation on such problem.

A more complicated alternative for estimating additive hazards regression is to pursue a Bayesian method proposed by Henschel et al. (2009) for right- and interval-censored data. Their estimation procedures are classified into two independent processes, of which one is used to calculate the nonparametric term, λ_0, and the other is responsible for estimating the vector of parametric term, β. For estimating λ_0, they adopt theories from spatial statistics (Rue, 2001) in which a sampling is conducted from Gaussian Markov Random Field (GMRF). As regarded with the parametric estimation, the likelihood function of Cox proportional hazards model is transformed into a generalized linear format with the log link so that the expectation of the form follows immediately a poisson distribution. Then they use Metropolis-Hasting algorithm for generalized linear model (Gamerman, 1997) incorporated with the values obtained from the estimation of λ_0 to provide the estimates of parametric term, β. After, they considered the situation where the frailty term(s) are present and suggest a solution for it.

Several other types of current status data are excluded from our discussions here and for their analyses, inference approaches which are able to yield special characteristics

of each type of the data are needed. These types include the current status data with time-dependent covariates (van der Laan and Robins, 1998), doubly censored current status data (Jewell and van der Laan, 1997, 2004a; Robinowitz and Jewell, 1996; van der Laan and Andrews, 2000; van der Laan et al., 1997; van der Laan and Jewell, 2001), and case-cohort current status data (Jewell and van der Laan, 2004b; Shiboski and Jewell, 1992). The doubly censored current status data usually refers to the situation while two consecutive events are present, only current status data are available. Case-cohort current status data, literally, means current status data obtained from case-cohort studies. Also in competing risk studies, one may face current status data (Jewell et al., 2003) that are generated from a cure survival model (Lam and Xue, 2005) involved in truncation and censoring simultaneously (Kim, 2003a).

As in any regression analysis, the criteria of selecting optimal model among other options can always attract special attentions from statistical scientists. In particular, this is also true during the process of modeling current status data with the underlying semi-parametric regressions. Within this area, many statistical procedures and diagnosis tools have been proposed, mostly for right-censored data (Klein and Moeschberger, 2003; Lawless, 2003), but there exists few methods specifically developed for current status data. One famous example is given by Ghosh (2003), where he discussed the goodness-of-fit of the additive hazards model (4.5) and developed some numerical and graphical methods based on the inference approach described in Chapter 4. Babineau (2005) also provided some illustrations about the goodness-of-fit for the fitting of parametric models to current status data. More detailed information on this is included in Sun (2006).

Appendix A

Bibliography

[1] Aalen, O. O. (1975). Statistical inference for a family of counting process. Ph.D. dissertation, Univ. of California, Berkeley.

[2] Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical models based on counting processes.* Springer-Verlag: New York.

[3] Aragon, J and Eberly, D. (1992). On convergence of convex minorant algorithm for distribution estimation with interval-censored data. *Journal of Computational and Graphical Statistics.* **1**, 129–140.

[4] Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972). *Statistical Inference under order restrictions.* New York: Wiley.

[5] Babineau, D. (2005). Goodness of fit for lifetime data models when responses are interval-censored. *Ph.D. Thesis*, University of Waterloo, Waterloo, Ontario, Canada.

[6] Breslow, N. E. and Day, N. E. (1987). *Statistical methods in cancer research*, **2**, *The design and analysis cohort studies.* Lyon:IARC

[7] Bertsekas, D. P. (1982). Projected newton methods for optimization problems with simple constraints. *SIAM Control and Optimization.* **20**, 221–246.

[8] Burr, D. (1994). A comparison of certain bootstrap confidence intervals in the cox model. *Journal of American Statistical Association.* **89**, 1290–1302.

[9] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B.* **34**, 187–220.

[10] Diamond, I. D. and McDonald, J. W. (1991). The analysis of current status data. *Demographic Applications of Event History Analysis*, eds. Trussel, J., Hankinson, R. and Tilton, J. Oxford University Press: Oxford, U.K.

[11] Diamond, I. D., McDonald, J. W. and Shah, I. H. (1986). Proportional hazards models for current status data: application to the study of differentials in age at weaning in Pakistan. *Demography.* **23**, 607–620.

[12] Dinse, G. E. and Lagakos, S. W. (1983). Regression analysis of tumor prevalence data. *Applied Statistics.* **32**, 236–248.

[13] Efron, B. and Tibshirani, R. J. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy (with discussion). *Biometrika.* **65**, 457–487.

[14] Finkelstein, D. M. and Wolfe, R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics.* **41**, 933-945.

[15] Fleming, T. R. and Harrington, D. P. (1991). *Counting process and survival analysis.* John Wiley: New York.

[16] Gamerman, D. (1997). Sampling from the posterior distribution in generalized linear mixed models. *Statistics for Computing 1997.* **7**, 57–68.

[17] Ghosh, D. (2001). Efficiency considerations in the additive hazards model with current status data. *Statistica Neerlandica.* **55**, 367–376.

[18] Ghosh, D. (2003). Goodness-of-fit methods for additive-risk models in tumorgenicity experiments. *Biometrics.* **59**, 721–726.

[19] Groeneboom, P. (1990). Nonparametric maximum likelihood estimation for interval censoring and deconvolution. Technical report 378. Stanford University, Statistics Department.

[20] Groeneboom, P. and Wellner, J. A. (1992). *Information bounds and non-parametric maximum likelihood estimation.* DMV Seminar, Band 19, Birkhauser, New York.

[21] Gross, S.T. and Huber-Carol, C. (1992). Regression models for truncated survival data. *Scandinavian Journal of Statistics.* **19**, 193–213.

[22] Greenwood M. (1926). The natural duration of cancer. *Reports on Public Health and Medical Subjects.* **33**, 1–26.

[23] Grummer-Strawn, L. M. (1993). Regression analysis of current status data: an application to breast-feeding. *Journal of the American Statistical Association.* **88**, 758–765.

[24] He, W. and Lawless, J. F. (2003). Flexible maximum likelihood methods for bivariate proportional hazards models. *Biometrics.* **59**, 837–848.

[25] Henschel, V., Engel, J., Holzel, D., and Mansmann, U. (2009). A semiparametric bayesian proportional hazards model for interval censored data with frailty effects. *Blood.* **21**, 699–716.

[26] Huang, J. (1995). Maximum likelihood estimation for proportional odds regression model with current status data. *Analysis of Censored Data.* IMS Lecture Notes - Monograph Series **27**, 129–146.

[27] Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *The Annals of Statistics.* **24**, 540–568.

[28] Huang, J. and Wellner, J. A. (1996). Regression models with interval-censoring. *Probability Theory and Mathematical Statistics (St. Petersburg, 1993).* 269–296.

[29] Huang, J. and Wellner, J. A. (1997). Interval censored survival data: a review of recent progress. *Proceedings of the First Seattle Symposium in Biostatistics*: Survival analysis, eds. Lin, D. and Fleming, T. Springer-Velbag, New York, 123–169 **21**, 699–716.

[30] Hoel, D. G. and Walberg, H. E. (1972). Statistical analysis of survival experiments. *Journal of National Cancer Institute.* **49**, 361–372.

[31] Jewell, N. P. and Shiboski, S. C. (1990). Statistical analysis of HIV infectivity based on partner studies. *Biometrics.* **46**, 1133–1150.

[32] Jewell, N. P. and van der Laan, M. J. (1996). Generalizations of current status data with applications. *Lifetime Data: models in Reliability and Survival Analysis*, Kluwer Acad. Publ., Dordrecht, 141-148.

[33] Jewell, N. P. and van der Laan, M. J. (1997). Singly and doubly censored current status data with extensions to multi-state counting processes. *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, eds. Lin, D. and Fleming, T. Springer-Verlag, New York, 171-184.

[34] Jewell, N. P., van der Laan, M. J. and Hennemean, T. (2003). Nonparametric estimation from current status data with competing risks. *Biometrika.* **90**, 183–197.

[35] Jewell, N. P. and van der Laan, M. J. (2004a). Current status data: review, recent developments and open problems. *Advances in Survival Analysis*, Elsevier, Amsterdam, 625-642.

[36] Jewell, N. P. and van der Lann, M. J. (2004b). Case-control current status data. *Biometrika*, **91**, 529–541.

[37] Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *Journal of Computational and Graphical Statistics.* **7**, 310–321.

[38] Kalbfleisch, J. D. and Lawless, J. F. (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association.* **80**, 863-871.

[39] Kalbfleisch, J. D. and Prentice, R. L. (2002). *The statistical analysis of failure time data.* Second edition, John Wiley: New York.

[40] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association.* **53**, 457–481.

[41] Keiding, N. (1991). Age-specific incidence and prevalence: A statistical perspective (with discussion). *Journal of the Royal Statistical Society, Series A.* **154**, 371–412.

[42] Keiding, N., Begtrup, K. Scheike, T. H. and Hasibeder, G. (1996). Estimation from current status data in continuous time. *Lifetime Data Analysis.* **2**, 119–129.

[43] Kim, J. and Lee, S. (1998). Two-sample goodness-of-fit tests for additive risk models with censored observations. *Biometrika.* **85**, 593–603.

[44] Kim, J. S. (2003a). Efficient estimation for the proportional hazards model with left-truncation and case-I interval-censored data. *Statistica Sinica.* **13**, 519–537.

[45] Klein, J. P. and Moeschberger, M. L. (2003). *Survival analysis.* Springer-Verlag: New York.

[46] Klein, R. W. and Spady, R. H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica.* bf 61, 387–421.

[47] Kulich, M. and Lin, D. Y. (2000). Additive hazards regression with covariate measurement error. *Journal of the American Statistical Association.* **95**, 238–248.

[48] Lagakos, S. W., Barraj, L. W., and Gruttola, V. (1988). Nonparametric analysis of truncated survival data, with application to AIDS. *Biometrika.* **75**, 515–523.

[49] Lawless, J. F. and Zhan, M. (1998). Analysis of interval-grouped recurrentevent data using piecewise constant rate functions. *The Canadian Journal of Statistics.* **26**, 549–565.

[50] Lam, K. F. and Xue, H. (2005). A semiparametric regression cure model with current status data. *Biometrika.* **92**, 573–586.

[51] Lin, D. Y., Oakes, D. and Ying, Z. (1998). Additive hazards regression with current status data. *Biometrika.* **85**, 289–298.

[52] Lin, D. Y. and Ying, Z. (1997). Additive hazards regression models for survival data. *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, eds. Lin, D. and Fleming, T. Springer-Verlag, New York, 185–198.

[53] Lin, D. Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika.* **81**, 61–71.

[54] Lin, D. Y. and Ying, Z. (1995). Semiparametric analysis of general additive multiplicative hazard models for counting processes. *The Annals of Statistics.* **23**, 1712–1734.

[55] Mangasarian, O. L. (1996). Algorithms of nonlinear mathematical programming. *CS730 Course Notes.* Univ. of Wisconsin, Computer Sciences Dept.

[56] Martinussen, T. and Scheike, T. H. (2002a). A flexible additive multiplicative hazard model. *Biometrika.* **89**, 283–298.

[57] Martinussen, T. and Scheike, T. H. (2002b). Efficient estimation in additive hazards regression with current status data. *Biometrika.* **89**, 649–658.

[58] Ng, M. P. (2002). A modification of Petos nonparametric estimation of survival curves for interval-censored data. *Biometrics.* **58**, 439–442.

[59] Ortega and Rheinholdt. (1970). The effect of 6-mercaptopmine on the duration of steroid induced remission in acute leukemia. *Analysis of Censored Data (Pune, 1994/1995)*, eds. H. L. Koul and J. V. Deshoande, IMS Lecture Notes, Monograph Series. **27**, 105–128.

[60] Pan, W. (1999). Extending the iterative convex minorant algorithm to the cox model for interval-censored data. *Journal of Computational and Graphical Statistics.* **8**, 109–120.

[61] Peto, R. (1973). Experimental survival curves for interval-censored data. *Applied Statistics.* **22**, 86–91.

[62] Polak, E. (1971). *Computational Methods in Optimization.* New York: Academic Press.

[63] Rabinowitz, D. and Jewell, N. P. (1996). Regression with doubly censored current status data. *Journal of the Royal Statistical Society, Series B.* **58**, 541–550.

[64] Reid, N. (1994). A conversation with Sir David Cox. *Statistical Science.* **9**, 439–455.

[65] Robertson, T. , Wright, F. T. and Dykstra, R. (1988). *Order restricted statistical inference.* John Wiley: New York.

[66] Rossini, A. J. and Tsiatis, A. A. (1996). A semiparametric proportional odds

regression model for the analysis of current status data. *Journal of the American Statistical Association.* **91**, 713–721.

[67] Rue, H. (2001). Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society (B) 2001.* **63**, 325–338.

[68] Schick, A. and Yu, Q. (2000). Consistency of the GMLE with mixed case interval-censored data. *Scandinavian Journal of Statistics.* **27**, 45–55.

[69] Scheike, T. H. and Zhang, M. (2002). An additive-multiplicative Cox-Aalen regression model. *Scandinavian Journal of Statistics.* **29**, 75–88.

[70] Self, S. G. and Grossman, E. A. (1986). Linear rank tests for interval-censored data with application to PCB levels in adipose tissue of transformer repair workers. *Biometrics.* **42**, 521–530.

[71] Shen, X. (2000). Linear regression with current status data. *Journal of the American Statistical Association.* **95**, 842–852.

[72] Shiboski, S. C. and Jewell, N. P. (1992). Statistical analysis of the time dependence of HIV infectivity based on partner study data. *Journal of the American Statistical Association.* **87**, 360–372.

[73] Shiboski, S. C. (1998). Generalized additive models for current status data. *Lifetime Data Analysis.* **4**, 29–50.

[74] Sun, J. (1998). Interval censoring. *Encyclopedia of Biostatistics.* John Wiley, First Edition, 2090–2095.

[75] Sun, J. (2005). Interval censoring. *Encyclopedia of Biostatistics.* John Wiley, Second Edition, 2603–2609.

[76] Sun, J. (2006). *The statistical analysis of interval-censored failure time data.* Springer-Verlag: New York.

[77] Sun,J. and Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *Journal of the Royal Statistical Society, Series B.* **62**, 293–302.

[78] Sun, L., Park, D. and Sun, J. (2006). The additive hazards model for recurrent gap times. *Statistica Sinica.* in press.

[79] Thall, P. F. (1988). Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics.* **44**, 197–209.

[80] Turnbull, B. W. (1976). The empirical distribution with arbitrarily grouped censored and truncated data. *Journal of the Royal Statistical Society, Series B.* **38**, 290–295.

[81] Thisted, R. A. (1988). *Elements of Statistical Computing.* New York: Chapman & Hall.

[82] van der Laan, M. J. and Robins, J. M. (1998). Locally efficient estimation with current status data and time-dependent covariates. *Journal of the American Statistical Association.* **93**, 693–701.

[83] van der Laan, M. J. and Andrews, C. (2000). The nonparametric maximum likelihood estimator in a class of doubly censored current status data models with application to partner studies. *Biometrika.* **87**, 61–71.

[84] van der Laan, M. J. Bickel, P. J. and Jewell, N. P. (1997). Singly and doubly censored current status data: estimation, asymptotics and regression. *Scandinavian Journal of Statistics.* **24**, 289–307.

[85] van der Laan, M. J. and Jewell, N. P. (2001). The NPMLE for doubly censored current status data. *Scandinavian Journal of Statistics.* **28**, 537–547.

[86] van der Laan,M. J. and Jewell, N. P. (2004b). Case-control current status data. *Biometrika.* **91**, 529–541.

[87] Wellner, J. A. (1995). Interval censoring case 2: alternative hypotheses. *Analysis of Censored Data (Pune, 1994/1995)*, eds. H. L. Koul and J. V. Deshoande, IMS Lecture Notes, Monograph Series **27**, 271–219.

[88] Wellner, J. A. and Zhan, Y. (1997). A hybird algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *Journal of the American Statistical Association.* **92**, 945–959.

[89] Xue, H., Lam, K. F. and Li, G. (2004). Sieve maximum likelihood estimation for semiparametric regression models with current status data. *Journal of the American Statistical Association.* **99**, 346–356.

[90] Zhan, Y. and Wellner, J. A. (1995). Double censoring: characterization and computation of the nonparametric maximum likelihood estimator. Technical Report. Univ. of Washington, Department of Statistics.

[91] Zhang, Z., Sun, J. and Sun, L. (2005). Statistical analysis of current status data with informative observation times. *Statistics in Medicine.* **24**, 1399–1407.