Analysis of Incomplete Event History Data

by

Min Zhan

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 1999

0-612-38288-5

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Abstract

Event history data arise in studies where a collection of individuals, each experiencing certain events or moving among a finite number of states, is followed over a period of time. The data consist of the number, time, type and sequence of events experienced by individuals, although the data are often incompletely observed. One example of such data comes from classical survival analysis, where individuals move from one state to the other, such as from alive to dead, or from healthy to diseased. The general event history data may contain information on events of multiple types or repeated occurrences of the same event (recurrent events).

The purpose of this thesis is to present methods using piecewise constant rate, intensity or hazard functions for event history data when events are interval-censored. These methods do not rely too heavily on parametric assumptions, and they are easier to implement than the nonparametric methods. In particular, we discuss the methods using piecewise constant rate, intensity or hazard functions for two types of event history data; one is interval-grouped recurrent events, the other is current status data and doubly-censored data.

Interval-grouped recurrent event data arise in longitudinal studies where subjects repeatedly experience a specific event and the events are observed only in the form of counts for intervals which can vary across subjects. We present two approaches for estimating the mean and rate functions of the recurrent event processes. One is mixed Poisson process estimation. Another is a robust method that requires only specification of the mean structure and covariance structure among recurrent event counts. Piecewise constant rate functions are incorporated in both

approaches. The two approaches are compared in a simulation study and in an example involving superficial bladder tumors in humans.

In many studies, interest focuses on the time between two successive events, the initiating event and the subsequent event. Current status data arise when the time of the initiating event is observed, but the only information for the subsequent event is whether it has occurred sometime between the initiating event and a single subsequent monitoring time. Doubly-censored data refer to data where both events are not observed directly, but are both interval-censored, or the initiating event is interval-censored and the subsequent event is right censored. We discuss methods with piecewise constant parametrization to estimate the survival function of the time between the two events for current status data and doubly-censored data. Different regression models are also developed. Simulation results show that our methods are robust to model misspecification. These methods are also applied to a data set from an AIDS study.

Finally, we explore the issue of getting smoother estimates of intensity, rate or hazard functions. A penalized likelihood approach is applied to the piecewise constant models. It is shown in a simulation study that this approach provides satisfactory estimates of the intensity, rate or hazard functions when events are interval-censored.

Acknowledgements

I am very grateful to my supervisor, Professor Jerry Lawless, for his guidance, encouragement and financial support during the process of my Ph.D. research. He is very patient and approachable when I need to discuss my research with him. I have benefitted greatly from his guidance.

I wish to thank Professors K.S. Brown, R.J. Cook, S.M. Smith and P.F. Thall, for their valuable comments and discussions.

Thanks also go to the Department of Statistics and Actuarial Science at University of Waterloo, for providing a nice environment for study.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Event History Data

Event history data arise in studies where a collection of individuals, each experiencing certain events or moving among a finite number of states, is followed over a period. The data consist of the number, time, type and sequence of events (*i.e.*, changes of states) experienced by individuals. This type of data is quite common in areas such as medicine, reliability, manufacturing, sociology, and demography. Some examples are: in clinical trials, the study of multi-type recurrent skin lesions (Abu-Libdeh *et al.*, 1990); in animal carcinogenicity experiments, the recurrence of tumors on rats (Lawless, 1987b); in the study of acquired immunodeficiency syndrome (AIDS), the estimation of the incubation period of AIDS (Bacchetti, 1990; Bacchetti and Jewell, 1991; Frydman, 1995); in manufacturing, the estimation of the rate of automobile warranty claims (Kalbfleisch *et al.*, 1991; Lawless and Nadeau, 1995).

The study of event history data is a fairly broad area (e.g., Andersen *et al.*, 1993). A well-known sub-field of this area is classical survival analysis, which involves individuals moving from one state to the other, such as from alive to dead, or from healthy to diseased. This thesis discusses more complicated event history data which may contain events of several types or recurrent events. We deal with so-called incomplete data. In particular, our main objective is to provide methods for situations where events are interval-censored, i.e., where the events are known only to lie in certain time intervals. Examples illustrating such data are given below.

1.1.1 Incidence of Nausea

The nausea data set, described by Thall and Lachin (1988), was from the National Cooperative Gallstone Study (NCGS). The NCGS was a multicenter, double-masked, placebo-controlled clinical trial. Patients with cholesterol gallstones were randomly assigned to one of the three groups: high-dose chenodiol, low-dose chenodiol or placebo. Patients were scheduled to visit their clinical center at successive follow-up dates and to report the number of episodes of nausea since their last visit. Since patients were often early, late, or missed scheduled visits, the actual visit times were irregular. One objective of the study was to assess the impact of treatment on the incidence of nausea.

The data set given by Thall and Lachin (1988) consists of the successive visit-times and the counts of episodes of nausea in each successive time interval during the first year of follow-up. The patients involved had floating gallstones, and were either in the high-dose chenodiol or in placebo groups. In this example, each patient could experience more than one occurrence of the same type of event. Moreover, the exact occurrence times of nausea were unknown; instead, only the counts in successive intervals were known.

1.1.2 Bladder Tumor Data

Byar (1980) discussed a randomized clinical trial for patients with bladder cancer; see also Sun and Wei (1996). All patients, who had superficial bladder tumors at the time of entering the trial, had their tumors removed and were assigned randomly to one of three treatments: placebo pills, pyridoxine, or thiotepa. At subsequent follow-up visits any tumors noticed were removed and the treatment was continued. The data consist of the months from the beginning of the study until each visit, the number of tumors present at each visit, the number of initial tumors at the time of randomization in the trial and the diameter of the largest of these tumors, for each patient. Once again, the exact times of tumor occurrence are not known.

1.1.3 A Rodent Tumorigenicity Experiment

Lindsey and Ryan (1993) presented a data set from a rodent tumorigenicity experiment conducted by the National Center for Toxicological Research. In this study, female mice were randomized to a control group or one of seven treatment groups, where the treatments are the known carcinogen 2-AAF at different dose levels . The onsets of tumors cannot be observed directly; instead, tumor presence can be detected only at the time of death or sacrifice. Hence the data set only consisted of numbers of deaths and sacrifices with and without tumors for each of five tumor types. The dose effect on tumor occurrence rate is one of the research interests. In this example, each mouse might experience several types of events, and the times of tumor onsets were only known to be in an interval.

1.1.4 Toronto Sexual Contact Study

The Toronto Sexual Contact Study (see Coates *et al.*, 1990; Yan and Lawless, 1992; Sun, 1995), conducted between 1984 and 1991, was a follow-up study of men infected with the human immunodeficiency virus(HIV). Two hundred and forty nine healthy homosexual or bisexual men who had at least one sexual contact with men diagnosed with AIDS were recruited into the study cohort between 1984 and 1985. Among them, 143 men were HIV positive at the time of recruitment, and 16 men seroconverted (ie, became HIV positive) during the study. We shall discuss only these 159 HIV positive subjects in this thesis. It was presumed that these men contracted the virus from their homosexual partners who had AIDS. For the subjects who were HIV positive at enrollment, the time of HIV infection was assumed to lie in the interval $[X_{Li}, X_{Ri}]$ defined by the dates of the subject's first and last sexual contacts with his sexual partner (index case). For the subjects who seroconverted during the study, the HIV infection time was considered to be observed and equal to the seroconversion time, although the actual HIV infection time was slightly earlier. Of the 159 subjects, only 49 were diagnosed with AIDS during the study and the times of AIDS diagnosis were known; the other subjects were AIDS-free by the date of last follow-up. The ages of subjects at the time of enrollment were recorded as a covariate.

In these examples, each subject can experience more than one occurrence of the same type of event, or each subject can experience several types of events. Moreover, the exact times at which the events occur may not be observed; that is, the data may be under some kind of interval-censoring, specified more precisely in Section 1.4. The purpose of this thesis is to develop relatively simple and appropriate methods for such event history data. We will mainly focus on two kinds of incomplete event history data; one is interval-grouped recurrent event data, such as in the incidence of nausea example, and the other is so-called doubly-censored data such as in the Toronto Sexual Contact Study example. We develop methodology based on piecewise-constant intensity, hazard or rate functions which is robust and rather easy to use. We will briefly review the literature on event history analysis in Section 1.2, introduce the methods we propose to develop for event history data in Section 1.3, and discuss some patterns of censoring and truncation in Section 1.4. Finally in Section 1.5, we will specify the objectives of the thesis and give the plan for the remaining chapters.

1.2 Review of Literature

The main objectives of event history analysis include the following:

1 (a) to estimate occurrence (incidence) rates of the events of interest;

or

1 (b) to estimate the conditional occurrence rates (intensities) of the events

given the previous history;

2 (a) to assess covariate effects on the occurrence rates of the events;

or

2 (b) to assess covariate effects on the conditional occurrence rates of the events given the previous history.

Methodology in classical survival analysis where there is a single event for each subject has been discussed by many authors (e.g., Kalbfleisch and Prentice, 1980; Lawless, 1982; Cox and Oakes, 1984). Approaches include Cox's partial likelihood analysis (Cox, 1972a) for the semiparametric proportional hazard models, parametric and nonparametric methods based on accelerated failure time models, and so on. In the following we will briefly review models for more general event history data, such as recurrent events or multiple types of events.

Let us consider a single subject. Suppose a type of event may repeatedly occur to this subject. Let $N(t)$ be the number of events occurred over time period $(0, t]$, and $T_1 < T_2 < \ldots$ be the occurrence times, where the T_j's are measured from the same time origin for this subject. The multiple types of events can be described by a multivariate process $\{N(t)\} = \{(N_1(t), \ldots, N_k(t))\}$, where $N_j(t)$ is the number of events of type j occurred for a subject up to time t. Assume the occurrence time sequence for the jth type of events is $T_{j1} < T_{j2} < \ldots$.

Models for event history data include intensity-based stochastic models and marginal models. We will discuss intensity-based stochastic models first, then marginal models. Finally we will mention frailty models which involve random effects.

1.2.1 Intensity-Based Stochastic Models

Full or complete models for event occurrence can be specified in terms of event intensity functions (Aalen, 1978; Fleming and Harrington, 1991; Andersen *et al.*, 1993). These specify the complete probabilistic structure of the event processes. The conditional intensity function for a simple type of recurrent event can be interpreted as the instantaneous occurrence rate of the event at time t, conditional on H_t which is the process history up to just before time t. That is,

$$\lambda(t; H_t) = \lim_{h \to 0} \frac{Pr\{N(t) - N(t-h) = 1 | H_t\}}{h}.$$

H_t may include previous events, covariate values, and possible censoring up to time t. In an intensity based model, $\lambda(t; H_t)$ is specified to be of some particular form. Let us consider the case of recurrent events.

A modulated nonhomogeneous Poisson process model with multiplicative intensity structure assumes that $\{N(t)\}$ is a Poisson process with intensity function $\lambda(t; H_t) = \lambda_0(t) g(z(t); \beta)$. Here $\lambda_0(t)$ is a nonnegative deterministic function, $z(t)$ may include the values of fixed and time dependent covariates, and g is a positive valued smooth function.

A modulated renewal process model (Cox, 1972b) on the other hand assumes that the intensity function is of the form $\lambda(t; H_t) = \lambda_0(t - t_{N(t-)}) g(z(t); \beta)$. If $z(t)$ is further assumed to be a time-independent covariate, the renewal process implies that the intensity function can be expressed in terms of inter-event times, since $t - t_{N(t-)}$ is the time since last event.

Aalen (1978); Fleming and Harrington (1991); Andersen *et al.* (1993) discussed multivariate counting processes with a multiplicative intensity assumption. Let $\{N(t)\}$ be a multivariate counting process, where $N(t) = (N_1(t), \ldots, N_k(t))$. The intensity function for the jth component is defined as

$$\lambda_j(t; H_t) = \lim_{h \to 0} \frac{Pr\{N_j(t) - N_j(t-h) = 1 | H_t\}}{h}.$$

The multiplicative intensity assumption states that $\lambda_j(t; H_t) = \alpha_j(t; Z(t))Y_j(t)$, where $\alpha_j(t; Z(t))$ is a nonnegative function depending on parameters and covariate values, and Y_j is a predictable function, which means the value of $Y_j(t)$ is fixed just before t; also assume Y_j does not involve parameters. Usually $Y_j(t)$ contains information about whether or not the subject is at risk for the jth event occurrence.

So-called multi-state models are used to describe situations where events for individuals correspond to changes of their "state" in life. A continuous time finite state Markov model (e.g., Ross, 1983; Andersen *et al.*, 1993) requires that for each individual, transitions between states follows a Markov process; *i.e.*, the conditional probability of transition from one state to another depends only on the current time and state occupied, not on the previous process history. Let $X(t)$ be the state occupied at time t by a given subject. The Markov process $\{X(t)\}$ satisfies, for arbitrary times $0 \leq s \leq t$,

$$Pr\{X(t) = j | X(s) = i, X(r), 0 \leq r < s\} = Pr\{X(t) = j | X(s) = i\}.$$

The transition intensity of moving from State i to State j at time t is defined as

$$q_{ij}(t) = \lim_{h\to 0} Pr\{X(t) = j|X(t-h) = i\}/h.$$

Definitions of the transition intensities may more generally include covariates. Continuous time Markov models have been widely applied to event history data (e.g., Andersen *et al.*, 1993).

A semi-Markov process model (e.g., Ross, 1983; Andersen *et al.*, 1993) requires that the conditional probability of a transition from one state to another at time t depends only on the current state occupied and the time since that state was entered. They are also widely used in some areas.

1.2.2 Marginal Models

Conditional or intensity-based models focus on the distribution of a process variable y_t at time t, conditioned on the process history H_t up to t. In contrast, marginal models focus on the distributions $f(y_t|z_t)$, where y_t is a response variable associated with the process at time t, and z_t is a covariate vector at time t. More generally one can consider responses (and covariates) associated with time intervals, for example the number of events in an interval, or outcomes at two or more specified points of time. Often the first two moments of the $y_t's$, that is , the mean $\mu_t = E(y_t|z_t)$ and covariance matrix $Var(y_t|z_t)$ conditional on covariates, are modeled (e.g., Thall, 1988; Thall and Vail, 1990; Lawless and Nadeau, 1995). Another approach is to model the marginal hazard function of specific event times (e.g., Wei *et al.*, 1989; Liang *et al.*, 1993; Lin, 1994). For a counting process, where $N(t)$ is the number of events occurred up to time t and $z(t)$ is an "external" covariate process that is not influenced by the event process $\{N(t)\}$, the mean function of the counting process is the expected number of events in the corresponding time interval, that is, $\mu(t|z(t)) = E\{N(t)|z(s), s \leq t\}$. The rate function $\rho(t|z(t))$ is defined as $\rho(t|z(t)) = d\mu(t|z(t))/dt$.

Ware *et al.* (1988) discussed the distinction between marginal models and conditional models in a multi-state framework related to a study about asthma symptoms. Lawless (1995) discussed marginal and conditional models for recurrent events.

1.2.3 Frailty Models

Observed covariates are often used to explain or model heterogeneity in event occurrences across different individuals. Another characteristic of event history data is the existence of unobserved individual heterogeneity, or inter-individual variation not explained by the observed covariates. This individual-level heterogeneity is often termed frailty in event history analysis. The sources of such heterogeneity include biological differences, unobserved or unrecorded environmental conditions, unobserved covariates, covariates measured subject to error, and so on. In the multivariate case, the latent variable or variables induce both correlation among the counts and extra-Poisson variation. A statistical model which includes random effects representing frailties is often called a frailty model in event history analysis. Usually the frailties are assumed to be time-constant, independent of covariates, and to act multiplicatively on the baseline intensity.

Frailty models have received considerable attention in the literature. They have been used to deal with the correlation among recurrent event times experienced by the same subject, or the association among event times in a subgroup of subjects, for example, siblings, or husbands and wives (Klein, 1992). Pickles and Crouchley (1994) gave a review of frailty models in survival and event history data. Aalen (1994) showed different ways to model frailty in survival analysis. Clayton (1994) discussed different methods for the analysis of recurrent event data and showed that frailty models can be seen as instances of generalized linear mixed models. Andersen *et al.* (1993) presented frailty model construction and maximum likelihood estimation for counting processes. Lawless (1987b) discussed frailty for Poisson models, and Aalen and Husebye (1991) and Follmann and Goldberg (1988) for renewal models.

Ignoring frailty effects often results in biased estimates of covariate effects, or biased estimates of risk of failure. Follmann and Goldberg (1988) gave an example where the ignorance of frailty resulted in a spurious decreasing hazard rate for observed failure times.

1.3 Our methods for event history data

The basic idea of our methods is to use a so called weakly parametric form, in particular, a piecewise constant form, for baseline intensity, rate or hazard functions. That is, there is a pre-specified sequence of constants $0 = a_0 < a_1 < \cdots < a_r < \infty$, such that a baseline event intensity (or rate, or hazard) $\lambda_0(t) = \rho_k$ for $t \in A_k = (a_{k-1}, a_k]$. Using a weakly parametric form for a baseline intensity avoids strong parametric assumptions and gives a reasonable approximation to the true baseline intensity function. Moreover, weakly parametric models avoid many problems associated with non-parametric methods for incomplete event history data.

Piecewise constant parametrization has been applied to the study of event history data by many authors; we mention a few related to our work. Lindsey and Ryan (1993) developed a three-state illness-death model using piecewise constant baseline transition intensity. Schluchter and Jackson (1989) considered a piecewise constant hazard model for right-censored survival data when covariates are categorical and partially observed. Carstensen (1996) discussed fitting of regression models with piecewise constant hazards to interval censored survival data, and implemented the models in standard statistical software. Kim (1997) discussed application of the EM algorithm to find maximum likelihood estimates for the parameters in a piecewise constant hazard model for interval-censored survival data. Hu and Lawless (1996) used piecewise constant models for recurrent events.

Under the piecewise constant intensity assumption, the estimated intensities may vary substantially between adjacent intervals when the number of pieces is large. A smoothed estimate of intensities can be obtained by maximizing the penalized log-likelihood function

$$log(L) - (1/2)\zeta J(\rho), \quad \zeta \geq 0,$$

where L is the likelihood function based on the data, J measures the roughness of the baseline intensity function, and ζ is a tuning constant. Penalized likelihood balances smoothness of the intensity function against the fit to the data. Green and

Silverman (1994) gave an overview of the roughness penalty approach (including penalized likelihood) to a wide range of smoothing problems. Examples of application of penalized likelihood approach in event history analysis include Bacchetti (1990), and Bacchetti and Jewell (1991), who discussed estimation of the incubation period of AIDS using penalized likelihood; and Fusaro *et al.* (1996), who discussed maximum penalized likelihood estimation of hazard functions.

1.4 Patterns of Censoring and Truncation

It is necessary to consider how the data on events for an individual are observed in order to provide valid inference. Since we have to stop observation at some time τ, or we may not start observation from the time origin of the event process, or we may not observe the event process continuously, some information about events may be missed. This creates "incomplete" data. Censoring and truncation are common ways to create "incomplete" data on events.

There is a considerable amount of discussion on censoring and truncation patterns in the literature (e.g., Aalen and Husebye, 1991; Andersen *et al.*, 1993). In the following we briefly discuss some patterns of censoring and truncation in event history data. These patterns are: right censoring, left censoring, interval censoring, right truncation, and left truncation.

If we stop observing the event process of an individual at time τ_1, we say that the event process is right-censored. If τ_1 is fixed in advance, or a random variable independent of the event process, or a random variable depending only on earlier observations of the event process, we say that τ_1 is a stopping time and the right-censoring is non-informative. The censoring time τ_1 can be treated as if it is fixed.

If we start observing the event process of an individual at time V that is larger than the time origin of the event process, we say the event process is left-censored. If a lifetime is left-censored, we know only that the lifetime is less than the time of starting observation. An inference based on the event process alone is valid only if V is a stopping time. The baboon descent data, given by Andersen *et al.* (1993), was an example of left censoring. Troops of baboons in Kenya sleep in the trees and descend for foraging at some time of the day. Observers often arrive later in the day than this descent and they can only know that descent happened before their arrival.

Interval-censoring refers to the case that the event process of an individual is only observed at discrete times $t_1 < t_2 < \ldots$ over some time interval $[\tau_0, \tau_1]$. In the incidence of nausea example in Section 1.1.1, the episodes of nausea are interval-censored.

Truncation is more generally a selection or sampling effect. With truncation, an individual is observed over a time interval $[\tau_0, \tau_1]$ conditional on some event A having occurred. If event A has not occurred, we would not observe this individual at all or we would not include this individual in the sample. In particular, the lifetime T_i for individual i is left-truncated if the condition for observing T_i is $T_i > V_i$, for some V_i. For example, in the Toronto Sexual Contact Study, the AIDS times were left-truncated since only men who had not been diagnosed with AIDS were allowed to enter into the study. Similarly, the lifetime T_i for individual i is right-truncated if the condition for observing T_i is $T_i < V_i$, for some V_i. For example, in a transfusion-related AIDS study, suppose an individual was infected with HIV at time T_1, developed AIDS at time T_2, and the closing date of the study was τ. Suppose also that only subjects who had AIDS by time τ were entered into the study. The incubation time $X = T_2 - T_1$ is then observed conditional on $X \leq \tau - T_1$. Kalbfleisch and Lawless (1989) discuss this setup in detail.

1.5 Plan of the thesis

In summary, the purpose of this thesis is to present methods using piecewise constant rate, intensity or hazard functions for event history data when events are interval-censored. In particular, we discuss the methods using piecewise constant rate, intensity or hazard functions for two types of event history data; one is interval-grouped recurrent events, the other is current status data and doubly-censored data.

Since we may not know the distributional form of the event times before analysis, we do not want to make a strong parametric assumption on the distribution. It is also desirable to have methods that are robust to the distributional form. The purely nonparametric or semiparametric methods satisfy this robustness requirement, but they are often hard to implement and standard errors of estimators are hard to obtain. The piecewise constant models can be considered as a compromise between the strongly parametric models and the purely nonparametric models. They are robust and relatively easy to implement. When the focus of our analysis is on the regression effect or the mean function of the event process, a piecewise constant model with a small number of pieces can be used; when we want to estimate the intensity, rate or hazard functions, a model with more pieces can be used with smoothing. The piecewise constant models can also serve as a tool for goodness of fit. By comparing the fit of a parametric model with the fit of a piecewise constant model, we can see how the parametric model fits to the data.

The plan for the rest of the thesis is as follows. Chapter 2 presents methods to estimate the mean functions for recurrent events and covariate effects based on mixed Poisson processes and estimating equations. Both methods use a piecewise constant intensity or rate function. These methods are compared in a simulation study and in an example involving superficial bladder tumors in humans. Chapter 3 considers analysis of current status data and doubly-censored data using piecewise constant hazard functions. The performance of the weakly parametric models is assessed in a simulation study and in an example from HIV/AIDS studies. Chapter 4 discusses the application of penalized likelihood techniques to produce smoothed estimates of intensity, rate or hazard functions for interval-grouped recurrent events and doubly-censored data. It is shown in a simulation study that this approach provides satisfactory estimates of the intensity, rate or hazard functions when events are interval-censored. Chapter 5 gives a summary for this thesis and discusses some related areas for further research.

Chapter 2

Analysis of Interval-Grouped Recurrent Event Data

2.1 Introduction

Studies in which individual subjects or units may experience recurrent events are common in many areas. For situations in which the exact times of event occurrence and covariates are observed there are well known methods of analysis based on point or counting processes (e.g., Andersen *et al.*, 1993; Lawless, 1995). However, Thall (1988), Thall and Lachin (1988) and others have discussed problems where only the numbers of events occurring in successive time intervals are known for each subject; moreover, the time intervals may vary from subject to subject. Thall and Lachin (1988) give an example where the recurrent events are episodes of nausea in a clinical trial for patients with gallstones. Another example is a bladder cancer study discussed by Byar (1980) in which superficial bladder tumors were observed and then removed at each visit to a clinic by a patient. Some authors refer to such data as panel count data, but we shall refer to them as interval-grouped recurrent events.

The problems may be discussed formally as follows. Let $N_i(t)$ denote the number of events occurring for subject i ($i = 1, \ldots, m$) over the continuous time interval $(0, t]$, and let z_i denote an associated $p \times 1$ vector of covariates. For subject i, we observe z_i and $n_{ij} = N_i(b_{ij}) - N_i(b_{i,j-1})$, the number of events in the interval $B_{ij} = (b_{i,j-1}, b_{ij}]$, $j = 1, \ldots, k_i$, where k_i is the number of intervals for which subject i is observed, and $0 = b_{i0} < b_{i1} < \cdots < b_{ik_i} = \tau_i$. The interval endpoints b_{ij} can be fixed or random, but they have to satisfy some conditions in order to make valid inference. These conditions are discussed in the next section. Our objective is to analyze the event occurrence processes, conditional on the z_i's.

Methods which do not rely too heavily on parametric assumptions or the precise nature of the event processes are frequently useful. Lawless and Nadeau (1995) have presented methodology for analyzing mean and rate functions, defined respectively by

$$\Lambda_i(t) = E\{N_i(t)\}, \quad \lambda_i(t) = \Lambda_i'(t). \tag{2.1}$$

They consider multiplicative specifications such as the log linear model where

$$\lambda_i(t) = \lambda_0(t) \exp(z_i'\beta), \tag{2.2}$$

where β is a $p \times 1$ regression parameter and $\lambda_0(t)$ is an arbitrary baseline rate function. Their methods assume only the validity of (2.1) and (2.2), and not that the recurrent events follow any specific type of process. However, they require that event times be observed.

Thall and Lachin (1988), Sun and Kalbfleisch (1993) and Sun and Kalbfleisch (1995) described nonparametric methods for estimating and comparing mean and rate functions in the case of interval-grouped data. However, their methods do not deal with general covariates and confidence interval estimation is problematic. Thall (1988) considered a parametric mixed Poisson model which deals with covariates. He adopted an intensity function with a form of $exp(\boldsymbol{f}_{ij}^T\boldsymbol{\beta}^{(1)} + \boldsymbol{X}_i^T\boldsymbol{\beta}^{(2)})$, where $\boldsymbol{f}_{ij}$ is a vector of known functions of time (no parameters are involved), and $\boldsymbol{X}_i$ is a vector of baseline covariates. He also approximated the expected count for a given observation interval by the product of the interval length and the intensity function value at the midpoint of the interval. Thall and Vail (1990) discussed a generalized estimating equation approach for longitudinal count data. They assumed that each expected count μ_{ij} depended on covariates $\boldsymbol{X}_{ij}$ and a vector of known functions of time $\boldsymbol{f}(\tau_{ij})$ through the form of $g(\mu_{ij}) = \boldsymbol{f}(\tau_{ij})^T\boldsymbol{\beta}^{(1)} + \boldsymbol{X}_{ij}^T\boldsymbol{\beta}^{(2)}$, while the covariance matrix $\boldsymbol{V}$ for counts $(n_{i1}, \cdots, n_{ik_i})$ was modelled parametrically, involving $\boldsymbol{\beta}^{(1)}$, $\boldsymbol{\beta}^{(2)}$ and some extra parameters. Staniswalis *et al.* (1997) extend Thall's (1988) approach by allowing $\lambda_0(t)$ to be nonparametric and employing a smoothing technique for its estimation. Sun and Matthews (1997) and Sun and Wei (1996) considered semiparametric estimation of regression parameters but not baseline rate functions.

The purpose of this chapter is to present methodology based on (2.1) and (2.2) for the general situation. We avoid strong parametric assumptions about $\lambda_0(t)$ and the need for complicated smoothing methods by using a piecewise constant form. This yields a different mean structure than those in Thall (1988) and Thall and Vail (1990). It is of course a theme of this thesis that the use of piecewise-constant intensity, rate and hazard functions avoids many problems associated with non- and semi-parametric methods for incomplete survival and event history data, while giving a high degree of robustness. Extensions to consider other forms than (2.2) for the regression specification are straightforward. We also avoid the Poisson process assumptions made by several authors by considering both mixed Poisson processes and robust methods. Finally, we investigate the robustness of the piecewise-constant formulation when the true underlying intensity is actually smooth, and compare two methods of estimating variance function parameters.

The remainder of this chapter is as follows. In Section 2.2 we develop methods for mixed Poisson processes with piecewise constant baseline rate functions. Although a specific process is assumed in this section, it is one which has been frequently found to be plausible. Section 2.3 presents robust methods that do not require that the event processes be mixed Poisson (or anything else). Section 2.4 presents a simulation study that assesses the performance and robustness of the methods of Section 2.2 and Section 2.3. Section 2.5 considers an example, and Section 2.6 concludes with some extensions of the methodology and suggestions for further work.

2.2 Mixed Poisson process estimation

We assume in this section that $\{N_i(t), t \geq 0\}$, $i = 1, \ldots, m$, are independent mixed Poisson processes (Lawless, 1987b; Thall, 1988). That is, conditional on α_i and covariate z_i, $\{N_i(t), t \geq 0\}$ is a nonhomogeneous Poisson process with intensity function

$$\lambda(t; \alpha_i, z_i) = \alpha_i \lambda_0(t) \exp(z_i' \beta), \tag{2.3}$$

where the α_i's are independent and identically distributed unobserved "frailty" variables with mean one and variance v; we will assume in this section that the α_i's follow a Gamma distribution. We take $\lambda_0(t)$ to be a piecewise constant baseline rate function, i.e., $\lambda_0(t) = \rho_k$ for $t \in A_k = (a_{k-1}, a_k]$, where $0 = a_0 < a_1 < \cdots < a_r < \infty$ is a pre-specified sequence of constants. Typically a model with r in the range of $4 - 10$ proves satisfactory. This range of r values gives flexible models with a fairly low-dimensional parameter. Our experience with the example in Section 5 and simulations has indicated that estimates of β, $\Lambda_0(t)$, the frailty variance parameter and their standard errors change very little once r is increased beyond 8 or 10. The mean function for this model is

$$\Lambda_i(t) = \Lambda_0(t) \exp(z_i' \beta), \tag{2.4}$$

where $\Lambda_0(t) = \sum_{k=1}^{r} \rho_k u_k(t)$, and $u_k(t) = \max(0, \min(a_k, t) - a_{k-1})$ is the length of the intersection of the interval $(0, t]$ with the interval A_k. Similarly, if we define $\mu_{ij} =$

$E\{N(b_{ij}) - N(b_{i,j-1})\}$, then $\mu_{ij} = \mu_{0,ij} \exp(z_i'\beta)$, where $\mu_{0,ij} = \sum_{k=1}^{r} \rho_k u_k(i,j)$, and

$$u_k(i,j) = \max(0, \min(a_k, b_{ij}) - \max(a_{k-1}, b_{ij-1}))$$

is the length of the intersection of interval B_{ij} with interval A_k.

To provide valid statistical analysis, we need to make assumptions concerning the relationship of the observation times and the event processes. The conditions derived by Gruger *et al.* (1991) are modified to meet our need here. We first define

$$H_{ij} = \{b_{i1}, n_{i1}, \ldots, b_{ij}, n_{ij}\}, \quad H_{i,j-} = \{b_{i1}, n_{i1}, \ldots, b_{ij}\},$$

and note that

$$\begin{aligned} P(H_{ik_i}|\alpha_i, z_i) &= \prod_{j=1}^{k_i} P(n_{ij}|H_{i,j-}, \alpha_i, z_i) \prod_{j=1}^{k_i} P(b_{ij}|H_{i,j-1}, \alpha_i, z_i) \\ &= L_{ia}L_{ib}. \end{aligned} \tag{2.5}$$

We also let $g(\alpha_i; v) = \alpha_i^{1/v-1} exp(-\alpha_i/v)/\{v^{1/v}\Gamma(1/v)\}$ be the density function of the Gamma distribution with mean 1 and variance v. Then the term $L_i = \int_0^\infty L_{ia} g(\alpha_i; v) d\alpha_i$ is the likelihood based on the data for the ith subject if the following two conditions are satisfied:

1. The probability of having n_{ij} occurrences of the event in the interval B_{ij}, given the history $H_{i,j-}$, is independent of the previous observation times, i.e.,

$$P(n_{ij}|H_{i,j-}, \alpha_i, z_i) = P(n_{ij}|n_{i1}, \ldots, n_{i,j-1}, \alpha_i, z_i).$$

2. $P(b_{ij}|H_{i,j-1}, \alpha_i, z_i)$ is independent of α_i and does not contain parameters of interest, i.e., (ρ, β, v), where $\rho = (\rho_1, \ldots, \rho_r)'$.

These conditions do not require that the b_{ij}'s be independent of the event processes. Instead, a conditional independence as in 1. is required; the current observation time b_{ij} can depend on previous observation times $b_{i1}, \ldots, b_{i,j-1}$ and previous counts $n_{i1}, \ldots, n_{i,j-1}$ but not on the "current" count n_{ij}. Thus, if a patient could anticipate that there was likely to be a large number of event occurrences during the current interval and went to the clinic for an early examination, conditions above would no longer hold.

If only the first condition is satisfied, L_i is a partial likelihood based on the ith subject, and inferences about the parameters may still be based upon it. However, we shall assume both conditions are satisfied and make inference about (ρ, β, v) based on the likelihood function $L = \prod_{i=1}^{m} L_i$.

Since

$$L_i = \int_0^\infty \prod_{j=1}^{k_i} \exp(-\alpha_i \mu_{ij})(\alpha_i \mu_{ij})^{n_{ij}} g(\alpha_i; v) d\alpha_i, \tag{2.6}$$

L can be simplified as

$$L \propto \prod_{i=1}^{m} \{\prod_{j=1}^{k_i} \mu_{ij}^{n_{ij}}\} \frac{\Gamma(n_{i.} + 1/v) v^{n_{i.}}}{\Gamma(1/v)(1 + v\mu_{i.})^{n_{i.}+1/v}}, \tag{2.7}$$

where $\mu_{i.} = \sum_{j=1}^{k_i} \mu_{ij}$ and $n_{i.} = \sum_{j=1}^{k_i} n_{ij}$. Therefore, the log-likelihood function is

$$\begin{aligned} l(\rho, \beta, v) &= \sum_{i=1}^{m} \sum_{j} n_{ij} \log \mu_{ij} + n_{i.} \log v + \log \Gamma(n_{i.} + 1/v) \\ &\quad - \log \Gamma(1/v) - (n_{i.} + 1/v) \log(1 + v\mu_{i.}), \end{aligned} \tag{2.8}$$

The likelihood score functions for (ρ, β, v) are

$$\begin{aligned}
\frac{\partial l}{\partial \rho_k} &= \sum_{i=1}^{m}\sum_{j=1}^{k_i}\frac{(n_{ij}-\mu_{ij})u_k(i,j)\exp(z_i'\beta)}{\mu_{ij}} - \\
&\quad \sum_{i=1}^{m}\frac{v(n_{i.}-\mu_{i.})u_k(i,+)\exp(z_i'\beta)}{1+v\mu_{i.}}, \qquad k=1,\ldots,r \qquad (2.9)\\
\frac{\partial l}{\partial \beta_s} &= \sum_{i=1}^{m}\frac{z_{is}(n_{i.}-\mu_{i.})}{1+v\mu_{i.}}, \qquad s=1,\ldots,p \qquad (2.10)\\
\frac{\partial l}{\partial v} &= \sum_{i=1}^{m}\{\frac{n_{i.}-\mu_{i.}}{v(1+v\mu_{i.})} + v^{-2}\log(1+v\mu_{i.})\} - \\
&\quad v^{-1}\sum_{i=1}^{m}\sum_{s=1}^{n_{i.}}[1+v(s-1)]^{-1}. \qquad (2.11)
\end{aligned}$$

The likelihood equations $\partial l/\partial \rho_k = 0$, $\partial l/\partial \beta_s = 0$, $\partial l/\partial v = 0$ can be solved by Newton's method or Fisher's scoring method. An alternative that is slower to converge but which avoids occasional divergence problems is the EM algorithm (Dempster *et al.*, 1977), which we now outline. Let n_{ijk} be the (unobserved) number of events in the interval $A_k \cap B_{ij}$ and consider the α_i's as data as well. Denote $\mu_{ijk} = \rho_k u_k(i,j)\exp(z_i'\beta)$. Then the full log-likelihood based on the α_i's and n_{ijk}'s is $l_{full}(\rho, \beta, v) = l_{full,1}(v) + l_{full,2}(\rho, \beta)$, where

$$\begin{aligned}
l_{full,1}(v) &= -m[\log\Gamma(1/v) + \frac{\log v}{v}] + \sum_i v^{-1}[\log\alpha_i - \alpha_i], \qquad (2.12)\\
l_{full,2}(\rho,\beta) &= \sum_{i=1}^{m}\sum_{j=1}^{k_i}\sum_{k=1}^{r} n_{ijk}\log\mu_{ijk} - \sum_{i=1}^{m}\alpha_i\mu_{i.}. \qquad (2.13)
\end{aligned}$$

Denote $\theta = (\rho^T, \beta^T, v)^T$. Given a current estimate $\theta^{(0)}$, the E-step computes

$$l_{E,1}(v) = E(l_{full,1}(v)|n_{ij}, j=1,\ldots,k_i, i=1,\ldots,m, \theta^{(0)}),$$

$$l_{E,2}(\rho,\beta) = E(l_{full,2}(\rho,\beta)|n_{ij}, j=1,\ldots,k_i, i=1,\ldots,m,\theta^{(0)}).$$

This gives (2.12) and (2.13) with the n_{ijk}'s, α_i's and $\log\alpha_i$'s replaced with estimates $\widetilde{n_{ijk}}$'s, $\widetilde{\alpha_i}$'s and $\widetilde{\log\alpha_i}$'s, which are defined as follows:

$$\widetilde{n_{ijk}} = n_{ij}\rho_k^{(0)}u_k(i,j)/\{\sum_{l=1}^{r}\rho_l^{(0)}u_l(i,j)\},$$

$\widetilde{\alpha_i} = C_{i1}/C_{i2}$, and $\widetilde{\log\alpha_i} = \psi(C_{i1}) - \log C_{i2}$, where $C_{i1} = n_{i.} + 1/v^{(0)}$, $C_{i2} = \mu_{i.}^{(0)} + 1/v^{(0)}$ and $\psi(t) = d\log\Gamma(t)/dt$. This yields

$$l_{E,1}(v) = -m[\log\Gamma(1/v) + \log v/v] + \sum_i v^{-1}[\widetilde{\log\alpha_i} - \widetilde{\alpha_i}], \tag{2.14}$$

$$l_{E,2}(\rho,\beta) = \sum_i\sum_j\sum_k \widetilde{n_{ijk}}\log(\mu_{ijk}) - \sum_i \widetilde{\alpha_i}\mu_{i..}. \tag{2.15}$$

The M-step maximizes (2.14) and (2.15) in v and (ρ,β) respectively. We iterate between the E-step and the M-step until convergence is achieved. We claim that convergence is obtained if two consecutive values of the parameters and log-likelihood differ very little. In detail, we stop iterations if

$$\max_{1\le k\le r+p+1} \frac{|\theta_k^{(1)} - \theta_k^{(0)}|}{|\theta_k^{(0)}| + 10^{-5}} \le \epsilon_1,$$

and

$$\frac{|l(\theta^{(1)}) - l(\theta^{(0)})|}{|l(\theta^{(0)})| + 10^{-5}} \le \epsilon_2,$$

where ϵ_1 and ϵ_2 are small positive numbers. We let $\epsilon_1 = \epsilon_2 = 10^{-4}$ in the procedures here.

The EM algorithm, like other general optimization algorithms, is not guaranteed to converge, and if it does, may not give the global maximum. See Dempster *et al.* (1977) and Wu (1983) for discussions. However, we can choose different starting values to see if they produce the same parameter estimates. If this is the case, we have some confidence that we have obtained the maximum. The speed of convergence for the EM algorithm depends on the shape of the likelihood, the number of parameters, starting values, the accuracy desired and other factors. With the convergence criteria above we found that very little computer time was needed to compute estimates. For a sample of size 85 and with recurrent counts varying from 1 to 38 it typically took about $150 - 200$ iterations to compute the estimates and their covariance matrix in a mixed Poisson process model with up to 10 pieces. Computation was programmed in FORTRAN.

When the maximum likelihood estimates $(\hat{\rho}, \hat{\beta}, \hat{v})$ are obtained, inferences can be based on the asymptotic distributions of likelihood ratio statistics or the asymptotic distributions of Wald statistics. The latter require second derivatives of the log likelihood (e.g., Lawless, 1987b), which are easily obtained. By noting that $E(-\partial^2 l/\partial v \partial \rho) = 0$ and $E(-\partial^2 l/\partial v \partial \beta) = 0$ we see that $(\hat{\rho}, \hat{\beta})$ and $\hat{v}$ are asymptotically independent.

2.3 Robust estimation

The methods in the previous section assume that the event processes are mixed Poisson. In this section, we present robust methods that do not need the Poisson assumption. These methods model the mean and the covariance matrix of the count responses within a subject. A group of generalized estimating equations is constructed and solved to get the estimates of the parameters. The assumptions about the observation process made in Section 2.2 are retained.

We assume $\mu_{ij} = E(n_{ij})$ is given by

$$\mu_{ij} = \mu_{0,ij} \exp(z_i'\beta), \tag{2.16}$$

where $\mu_{0,ij} = \sum_{k=1}^{r} \rho_k u_k(i,j)$ as defined in the previous section. The 'working' covariance matrix for $\boldsymbol{n}_i = (n_{i1}, \cdots, n_{ik_i})^T$ will have the following form:

$$\boldsymbol{V}_i = cov(\boldsymbol{n}_i) = \boldsymbol{C}_i + v\boldsymbol{\mu}_i\boldsymbol{\mu}_i^T, \tag{2.17}$$

where $\boldsymbol{C}_i = diag(\boldsymbol{\mu}_i)$ and $\boldsymbol{\mu}_i = (\mu_{i1}, \cdots, \mu_{ik_i})^T$. This is the covariance matrix for the mixed Poisson model of Section 2.2, but the methods below are robust to the form of $\boldsymbol{V}_i$. Other forms that are plausible for specific applications can equally well be used.

Letting $\boldsymbol{D}_i = \partial\boldsymbol{\mu}_i/\partial(\boldsymbol{\rho}^T, \boldsymbol{\beta}^T)$, we define the generalized estimating equations for $\boldsymbol{\rho}$ and $\boldsymbol{\beta}$ as follows (c.f. McCullagh and Nelder 1989, Chap.9):

$$\boldsymbol{U}_1 = \sum_{i=1}^{m} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{S}_i = 0, \tag{2.18}$$

where $\boldsymbol{S}_i = \boldsymbol{n}_i - \boldsymbol{\mu}_i$. These give two subsets of equations:

$$\boldsymbol{U}_1 = \begin{pmatrix} \sum_{i=1}^{m} \boldsymbol{U}_{11i} \\ \sum_{i=1}^{m} \boldsymbol{U}_{12i} \end{pmatrix}, \tag{2.19}$$

where

$$U_{11i} = \exp(z_i'\beta)\left(\sum_j \frac{(n_{ij}-\mu_{ij})}{\mu_{ij}}u(i,j) - \frac{v(n_{i.}-\mu_{i.})}{1+v\mu_{i.}}u(i,+)\right) \quad (2.20)$$

$$U_{12i} = \frac{(n_{i.}-\mu_{i.})}{1+v\mu_{i.}}z_i, \quad (2.21)$$

where $u(i,j) = (u_1(i,j), \cdots, u_r(i,j))^T$, and $u(i,+) = \sum_{j=1}^{k_i} u(i,j)$.

One way to estimate ρ and β is to adopt working covariance matrices V_i based on a specified value of v. In particular, one could use the value $v = 0$, in which case V_i is the covariance matrix from a Poisson process for the events. The equations (2.18) are readily solved, for example by Newton's method, to give estimates $\tilde{\rho}$ and $\tilde{\beta}$. It follows from standard results for estimating equations (e.g., White, 1982; Breslow, 1990) that under mild conditions on the event processes and the observational scheme, the estimators $\tilde{\rho}$ and $\tilde{\beta}$ and other estimators below are consistent and asymptotically normally distributed as $m \to \infty$. The types of conditions needed are discussed in references such as Kaufmann (1987) and McCullagh and Nelder (1989). In particular, the observation scheme should ensure that the eigenvalues of the information matrix analogue $mG_{m,11}$ in (2.26) increase without limit as $m \to \infty$. Inferences about ρ or β can be based on the asymptotic normal distribution for $m^{1/2}\left((\tilde{\rho}-\rho)^T, (\tilde{\beta}-\beta)^T\right)^T$, whose covariance matrix may be estimated consistently by (2.30).

Alternatively, we may choose to use working covariance matrices which include one or more unknown dispersion parameters v. To estimate v in (2.17), we propose to use the following moment equation:

$$U_2 = \sum_{i=1}^{m} w_i\{(n_{i.} - \mu_{i.})^2 - \sigma_{i.}^2\} = 0, \tag{2.22}$$

where $\sigma_{i.}^2 = var(n_{i.}) = \mu_{i.} + v\mu_{i.}^2$, and the w_i are selected weights. We consider the choices (a) $w_i = 1/\sigma_{i.}^2$, (b) $w_i = 1$, (c) $w_i = \mu_{i.}^2/\sigma_{i.}^4$. They correspond to weights used in estimating equations for mixed Poisson regression models (Dean, 1991).

The estimates $\tilde{\rho}$, $\tilde{\beta}$ and $\tilde{v}$ can be obtained by iterating between (2.20), (2.21) and (2.22). We first update (ρ, β) by a modified Fisher scoring approach:

$$\begin{aligned}\left(\rho^{(1)T}, \beta^{(1)T}\right)^T &= \left(\rho^{(0)T}, \beta^{(0)T}\right)^T + \{\sum_{i=1}^{m} D_i^T V_i^{-1}(\rho^{(0)}, \beta^{(0)}, v^{(0)}) D_i\}^{-1} \\ &\quad \{\sum_{i=1}^{m} D_i^T V_i^{-1}(\rho^{(0)}, \beta^{(0)}, v^{(0)}) S_i(\rho^{(0)}, \beta^{(0)})\},\end{aligned} \tag{2.23}$$

then v is updated by

$$v^{(1)} = v^{(0)} - \frac{\partial U_2}{\partial v}\Big|^{-1}_{\rho^{(1)}, \beta^{(1)}, v^{(0)}} U_2(\rho^{(1)}, \beta^{(1)}, v^{(0)}). \tag{2.24}$$

We iterate between the two steps until (ρ, β, v) converges.

If the V_is are correct, $\tilde{\rho}$, $\tilde{\beta}$ and $\tilde{v}$ are consistent and asymptotically normal as $m \to \infty$, under an extension of conditions mentioned above. The asymptotic covariance matrix of $m^{1/2}(\tilde{\rho}^T - \rho^T, \tilde{\beta}^T - \beta^T, \tilde{v} - v)^T$ can be consistently estimated by

$$G_m^{-1}(\tilde{\rho}, \tilde{\beta}, \tilde{v}) \widetilde{H}_m G_m^{-T}(\tilde{\rho}, \tilde{\beta}, \tilde{v}), \tag{2.25}$$

where

$$G_m = \begin{pmatrix} G_{m,11} & G_{m,12} \\ G_{m,21} & G_{m,22} \end{pmatrix}, \tag{2.26}$$

with

$$\begin{aligned}
G_{m,11} &= m^{-1}\sum_i D_i^T V_i^{-1} D_i = m^{-1}E\{-\partial U_1/\partial(\rho^T,\beta^T)\}, \\
G_{m,12} &= 0 = m^{-1}E\{-\partial U_1/\partial v,\} \\
G_{m,21} &= m^{-1}\sum_{i=1}^{m} w_i(1+2v\mu_{i.})\frac{\partial \mu_{i.}}{\partial(\rho^T,\beta^T)} = m^{-1}E\{-\partial U_2/\partial(\rho^T,\beta^T)\}, \\
G_{m,22} &= m^{-1}\sum_{i=1}^{m} w_i\mu_{i.}^2 = m^{-1}E\{-\partial U_2/\partial v\}, \\
H_m &= m^{-1}cov(U),
\end{aligned} \tag{2.27}$$

and H_m is estimated by

$$\widetilde{H}_m = \begin{pmatrix} \widetilde{H}_{m,11} & \widetilde{H}_{m,12} \\ \widetilde{H}_{m,21} & \widetilde{H}_{m,22} \end{pmatrix}, \tag{2.28}$$

with

$$\begin{aligned}
\widetilde{H}_{m,11} &= m^{-1}\sum_i (D_i^T V_i^{-1} S_i S_i^T V_i^{-1} D_i), \\
\widetilde{H}_{m,12} &= m^{-1}\sum_i D_i^T V_i^{-1} S_i w_i[(n_{i.}-\mu_{i.})^2-\sigma_{i.}^2], \\
\widetilde{H}_{m,21} &= \widetilde{H}_{m,12}^T, \\
\widetilde{H}_{m,22} &= m^{-1}\sum_i w_i^2[(n_{i.}-\mu_{i.})^2-\sigma_{i.}^2]^2.
\end{aligned} \tag{2.29}$$

These matrices are evaluated at $(\tilde{\rho},\tilde{\beta},\tilde{v})$.

When the V_is are correct, the asymptotic covariance matrix of $m^{1/2}(\tilde{\rho}^T - \rho^T, \tilde{\beta}^T - \beta^T)^T$ is more efficiently estimated by $G_{m,11}^{-1}(\tilde{\rho}, \tilde{\beta}, \tilde{v})\widehat{H}_{m,11}G_{m,11}^{-T}(\tilde{\rho}, \tilde{\beta}, \tilde{v})$, where $\widehat{H}_{m,11} = H_{m,11}(\tilde{\rho}, \tilde{\beta}, \tilde{v})$, and H_m is partitioned in the same fashion as G_m. Since $H_{m,11} = G_{m,11}$ when the V_is are correct, the variance estimate reduces to $G_{m,11}^{-1}(\tilde{\rho}, \tilde{\beta}, \tilde{v})$. When the V_is are incorrect, $\tilde{\rho}$ and $\tilde{\beta}$ are still consistent and asymptotically normal. The asymptotic covariance matrix can then be consistently estimated by (e.g., Liang and Zeger, 1986)

$$G_{m,11}^{-1}(\tilde{\rho}, \tilde{\beta}, \tilde{v})\widetilde{H}_{m,11}G_{m,11}^{-T}(\tilde{\rho}, \tilde{\beta}, \tilde{v}), \tag{2.30}$$

which is the submatrix of (2.25) pertaining to $(\tilde{\rho}, \tilde{\beta})$. To guard against misspecification of V_i, this estimate is recommended for general use.

Confidence intervals or tests about parameters or functions thereof can be based on the approximate normality of $(\tilde{\rho}^T, \tilde{\beta}^T, \tilde{v})^T$ when samples are sufficiently large. In the next section we provide some simulation results to assess the performance of these methods.

2.4 Simulation study

A simulation study was carried out to assess the performance of the methods of Section 2.2 and 2.3. In particular, we wanted to (i) see how well the piecewise-constant rate functions allows us to estimate a continuous, smooth mean function, (ii) how well the estimating equation (GEE) methods of Section 2.3 performed compared to maximum likelihood methods, and (iii) how well normal approximations and covariance estimates used to obtain confidence intervals perform. With this in mind we generated interval-grouped recurrent event data from a mixed Poisson process with a "Weibull" baseline rate function $\lambda_0(t) = \gamma\delta t^{\delta-1}$ in the following way:

1. Generate independent Gamma variables $\alpha_1, \ldots, \alpha_m$ with mean 1 and variance v;

2. For given α_i, covariate $z_i = (z_{i1}, \ldots, z_{ip})$, regression coefficient β, and grouping interval endpoints b_{ij}'s, generate independent Poisson variables n_{ij} with means

$$\mu_{ij} = \alpha_i \exp(z_i'\beta)\{\Lambda_0(b_{ij}) - \Lambda_0(b_{i,j-1})\}, \quad j = 1, \ldots, k_i, \quad i = 1, \ldots, m$$

   where $\Lambda_0(t) = \gamma t^\delta$.

We selected sample size and the values of parameters and covariates to reflect situations commonly encountered in practice. However, only the case of 1 covariate (p=1) was considered. The simulation settings were as follows:

(1) m=90;

(2) One third of the z_i's were each of $-1, 0, 1$;

(3) Sequences of observation times b_{ij}'s are preset and do not depend on covariate values. For the thirty subjects with $z_i = -1$, we let the first 6 subjects be observed at times $t = 1, 4, 7, 12, 18$; the next 15 subjects are observed at $t = 2, 5, 9, 14, 21, 28, 35$; the last 9 subjects are observed at $t = 1, 3, 8, 14, 20, 26, 32, 38, 44, 50$. We set the sequences of observation times among subjects with covariate values 0 or 1 the same way.

(4) Eight different settings of the other parameters were considered. They are:

(a) $\gamma = 0.8, \delta = 0.5, \beta = 1.5, v = 0.5$; (b) $\gamma = 0.8, \delta = 0.5, \beta = 1.5, v = 0.2$;

(c) $\gamma = 0.8, \delta = 1.0, \beta = 1.5, v = 0.5$; (d) $\gamma = 0.8, \delta = 1.0, \beta = 1.5, v = 0.2$;

(e) $\gamma = 0.8, \delta = 0.5, \beta = 0.375, v = 0.5$; (f) $\gamma = 0.8, \delta = 0.5, \beta = 0.375, v = 0.2$;

(g) $\gamma = 0.8, \delta = 1.0, \beta = 0.375, v = 0.5$; (h) $\gamma = 0.8, \delta = 1.0, \beta = 0.375, v = 0.2$.

These parameter values generate values of $\mu_{i.}$ that range from a low of .76 to a high of 179. For each setting the values of $\mu_{i.}$ for the 9 combinations $\{-1, 0, 1\} \times \{18, 35, 50\}$ of z_i and τ_i are as follows:

(a) and (b). 0.76, 1.06, 1.26, 3.39, 4.73.5.66, 15.2, 21.2, 25.4;

(c) and (d). 3.21, 6.25, 8.93, 14.4, 28, 40, 64.5, 125, 179;

(e) and (f). 2.33, 3.25, 3.39, 3.89, 4.73, 4.94, 5.66, 6.89, 8.23;

(g) and (h). 9.90, 14.4, 19.2, 21.0, 27.5, 28, 40, 40.7, 58.2.

For each setting we generated 100 samples. The parameters β, v, and $\Lambda_0(t)$, were estimated by four approaches: (I) maximum likelihood of Section 2 based on a mixed Poisson process (gamma-Poisson) with a piecewise-constant rate function, (II) the GEE approach of Section 3 with a piecewise-constant rate function, (III) maximum likelihood based on a mixed Poisson process with a Weibull rate function, (IV) the GEE approach with a Weibull rate function. The cut points a_k used in (I) and (II) were $0, 5, 10, 15, 20, 25, 30, 40, 50$, so there were 8 pieces in the rate function. For methods II and IV two estimating equations for v were considered: (2.22) with (1) $w_i = \mu_{i.}^2/\sigma_{i.}^4$ or (2) $w_i = 1/\sigma_{i.}^2$. This gave us a chance to compare estimating functions for variance parameters, as in Dean (1991).

There is very little difference in the averages across the estimation methods, and all indicate little bias. Tables 2.1-2.4 show average values of $\widehat{\Lambda_0(25)}$, $\widehat{\Lambda_0(48)}$, $\hat{\beta}$ and $\hat{v}$ for settings (a), (b), (g) and (h); results for the other settings were similar. Here "$\wedge$" denotes either maximum likelihood or robust estimates, as indicated. The adoption of a piecewise constant baseline rate function also gives a close approximation to the maximum likelihood estimate based on the correct smooth model, as far as estimation of the mean function $\Lambda_0(t)$ is concerned. For example, Figure 2.1 shows the average of estimates $\hat{\Lambda}_0(t)$ based on the piecewise constant and smooth rate functions for case (a), along with the true mean function. The piecewise-constant model differs only in the time interval $(0, 5)$; this is due to the high true rate near $t = 0$, and could be alleviated by splitting the interval $(0, 5)$ into 2 or 3 pieces.

Tables 2.1-2.4 also present the empirical standard errors of $\widehat{\Lambda_0(25)}$, $\widehat{\Lambda_0(48)}$, $\hat{\beta}$, $\hat{v}$, and their average standard errors based on asymptotic theory estimates. For maximum likelihood approaches, standard errors are computed from the inverse of the expected information matrix; for the GEE approaches, standard errors are computed from the 'sandwich' type variance estimates (2.25). The approaches using piecewise-constant rates (I,II-1,II-2) and approaches using the true Weibull rate function (Approach III, IV-1, IV-2) have little difference in the averages of $\hat{\beta}$, $\hat{v}$, se($\hat{\beta}$) and se($\hat{v}$); they also produce similar estimates for $\Lambda_0(t)$ when t is not too small. Approaches III, IV-1,IV-2 are only slightly more efficient for $\Lambda_0(t)$.

The averages of the standard errors based on asymptotic theory are quite close to the empirical standard errors of the estimates. There is also no difference in the performance of the two methods of estimation of the variance parameter v (methods

1 and 2 in the tables); they correspond to weight functions (c) and (a), respectively, following (2.22). We note for interest that weight function (b) did not agree quite as well.

We also examined the empirical coverage of 90% and 95% confidence intervals for $\Lambda_0(25)$, $\Lambda_0(48)$, β and v based on normal approximations. Bearing in mind the small number of simulation runs (100), there were no major discrepancies between actual and nominal coverage for $\Lambda_0(25)$, $\Lambda_0(48)$ and β. There were discrepancies for v when the GEE approach with variance estimate (2.25) was used. Coverage of confidence intervals for v based on the GEE approach are improved slightly by using normal approximations for $\log(\hat{v})$ rather than $\hat{v}$. Table 2.5 and 2.6 show sample results for parameter settings (a) and (g). However, as has been demonstrated for the simpler but closely related case of mixed Poisson regression, confidence interval estimation of v is problematic in many situations and deserves further study; see Section 4 of Lawless (1987a). Tests of the null hypothesis $v = 0$ are also of interest, and equally deserve further study. Ng and Cook (1997) discuss score tests developed by them and earlier authors for the mixed Poisson process case. As far as estimation of the regression coefficients and baseline mean function are concerned, the difficulty of confidence interval estimation of v has little effect in most practical situations.

To summarize, the use of a piece-wise constant baseline rate function provides excellent estimation of regression coefficients and mean functions and, for scenarios similar to those in the simulations, confidence intervals based on normal approximations perform well. It seems likely that these properties will hold for other scenarios in which there are moderately many subjects (say 50 or more) and where not too many subjects have very small expected counts. In other situations, additional checks by simulation are recommended.

2.5 An example

The bladder cancer data (see, Byar, 1980) are described in Chapter 1. The data are interval-censored recurrent event data. Details about how the b_{ij}'s were determined are not given in the references cited. We assume that they satisfy conditions 1 and 2 of Section 2.2.

In the following analysis, we consider only patients in the placebo or thiotepa groups. There are 47 and 38 patients in the placebo group and thiotepa group, respectively. The time in study for each patient ranged from one month to 53 months, and the number of visits for each patient ranged from one to 38. Figure 2.2 gives a scatter plot of the total number of tumors present against the time in study for each patient. We notice that there is a fairly large number of patients with no tumor present.

Mixed Poisson processes with piecewise-constant rates were fitted to the data. We divided the whole study period $(0, 53]$ into 8 pieces, with the cut points

$$0, 5.5, 10.5, 15.5, 20.5, 25.5, 30.5, 40.5, 53.$$

This made the first 6 intervals of nearly equal length and the last two longer, since there were fewer subjects at risk for the last two intervals. We define the following variables: for the ith patient, $z_{i1} = 1$, if the patient is in the thiotepa group, $z_{i1} = 0$, otherwise; z_{i2} is the number of initial tumors present at randomization; z_{i3} is the diameter (in centimeters) of the largest initial tumor. Table 2.7 lists the MLEs and the standard errors based on the asymptotic distributions of the MLEs, as discussed in Section 2.

The robust estimation procedure of Section 3 was also applied to this data set. Table 2.7 also lists the robust estimates and their standard errors based on the robust 'sandwich' type variance estimates (2.25). We used $w_i = \mu_{i.}^2/\sigma_{i.}^4$ in U_2 (2.22).

We notice that the mixed Poisson process models and the generalized estimating equation approach produce similar estimates for ρ and β. Based on either approach, we conclude that patients in the thiotepa group have lower rate of tumor recurrence than patients in the placebo group, and patients with a large number of initial tumors have higher rate of tumor recurrence if all other conditions are the same. The estimates of v and their standard errors differ somewhat. As has been noted by Breslow (1990) and others, the maximum likelihood estimation of v is not robust to departures from the assumed mixed Poisson model (here, gamma-Poisson), whereas the robust procedures are, assuming satisfactory specification of the rate function. Consequently we prefer the robust estimate of v, and standard errors for the other parameters that are based on it.

Residual plots based on the $n_{i.}$'s were constructed to check the fit of our models. 'Anscombe' residuals were used, defined as

$$e_i = \frac{3(n_{i.}^{2/3} - \hat{\mu}_{i.}^{2/3})}{2\hat{\mu}_{i.}^{1/6}(1 + \hat{v}\hat{\mu}_{i.})^{1/2}}, \quad (2.31)$$

where $n_{i.}$ and $\hat{\mu}_i$ are the observed and the estimated numbers of tumors present for Patient i, respectively. The justification for using e_i was that the Anscombe residual is less skewed than the Pearson residual for Poisson variables, and should retain this property for negative binomial variables. Figure 2.3 shows the plot of e_i against the estimated number of tumors $\widehat{\mu}_{i.}$ based on the mixed Poisson model. We notice that all e_i's are between -1 and 2 except one value; there is no clear pattern except the following: there is one isolated point (corresponding to patient #16) with a large estimated number of tumors; moreover, a curve is observed at the left bottom of the plot, which corresponds to the observations with $n_{i.} = 0$. Plots of e_i against covariates z_1 and z_2 were also examined. None of the plots indicates major problems with the model. A more detailed assessment of fit based on the interval counts for each subject also does not reveal major problems, though the sparseness of the data makes formal tests difficult.

We also fitted models to the data after deleting observation #16. There are some changes in the estimates of parameters and their standard errors, but the significance level for β does not change much. Finally, in view of the large number of subjects (38) with $n_{i.} = 0$, we computed the estimated frequency of patients with $n_{i.} = 0$; it is 35.6 under the mixed Poisson model, and does not indicate any lack of fit.

2.6 Concluding remarks

Models with piecewise-constant intensities, rates or hazard functions provide an attractive approach in problems involving failure times or recurrent events when data are interval-censored, double-censored, or truncated in some way. In those situations purely non- or semi-parametric methods are often difficult to implement or to use for purposes of testing or interval estimation (e.g., Sun and Kalbfleisch, 1993, 1995; Jewell *et al.*, 1994). In this chapter we have examined the case of interval-grouped recurrent events in some detail.

The simulations in Section 2.4 were done under the assumption of mixed Poisson processes, and the robust methods performed essentially as well as maximum likelihood. In fact, for the negative binomial model of Section 2.2, which was used for the simulations, the maximum likelihood estimating functions (2.9) and (2.10) for ρ and β are the same as the robust estimating functions from (2.20) and (2.21), and are valid beyond mixed Poisson processes. However, variance estimation and confidence intervals based on maximum likelihood are non-robust, and consequently we recommend using the robust procedures if there is any doubt as to the underlying processes. These observations and recommendations are similar to those of Lawless (1987a) and Breslow (1990) in the context of ordinary regression analysis of counts.

The robust methods of Section 2.3 and the Poisson process-based methods of Section 2.2 are both readily extended to deal with other problems. For example, Hu and Lawless (1996) consider the estimation of rate and mean functions from zero-truncated recurrent event data; the methods here could be adapted to the case where the data are also interval-grouped. Similarly, discrete mixture models such as the ZAP or ZIP models of Lambert (1992) and Heilbron (1989) can be handled rather easily; these authors used Poisson process assumptions which are often best avoided because of the likelihood of overdispersion. It should be noted that in cases involving selection or truncation, such as when only individuals with at least one event are observed (Hu and Lawless, 1996), it is straightforward to develop methods based on Poisson mixtures as in Section 2.2, but usually difficult to develop methods analogous to those in Section 2.3.

In most practical situations it is satisfactory to use piecewise-constant intensities with $4-10$ pieces. However, where smoother rate function estimates are desired, an alternative is to increase the number of pieces substantially and use penalized likelihood approach (e.g., Bacchetti, 1990; Green and Silverman, 1994). This approach is studied in Chapter 4. Other approaches include kernel functions (Staniswalis *et al.*, 1997) and B-splines (Rosenberg, 1995), which give smooth rate functions. Staniswalis *et al.* (1997) examine a kernel smoothing approach in a semiparametric regression model for the applications in this chapter. They apply the generalized profile likelihood method of Severini and Wong (1992) to get the estimates of the baseline rate function and covariate effects. Their procedure provides desirable estimates of the hazard function, but it is computationally intensive. Rosenberg (1995) models the baseline hazard as a linear combination of cubic B-splines. In principle, this approach is very similar to the piecewise constant hazard approach because both are weakly parametric models. As the number of knots or cut-points increases, both the spline estimates and the smoothed piecewise constant estimates of the intensity functions would have similar values.

In our approaches, we assume the number and location of the cut-points are pre-specified. An alternative way is to allow them to be selected based on data.

Rosenberg (1995) proposes choosing the number and location of the knots in the B-spline approach based on Akaike's information criterion (Akaike, 1973). It can also be used to choose the number and location of the cut-points in the piecewise constant hazard models.

We have seen that our methods work well for the simulated data sets with $m = 90$, one covariate and eight pieces in the piecewise constant models. In the example of bladder cancer data, we have also seen that our methods can handle the case of three covariates easily with $m = 85$. We anticipate that a data set with a larger sample size and a vector of 4 or 5 covariates will not bring significant numerical problems.

Table 2.1: The sample average of estimates and estimated standard errors based on asymptotic theory, and the empirical standard errors, in setting (a) $\gamma = 0.8$, $\delta = 0.5$, $\beta = 1.5$, $v = 0.5$, $\Lambda_0(25) = 4$, $\Lambda_0(48) = 5.543$

	I	II-1	II-2	III	IV-1	IV-2
$\widehat{\Lambda_0(25)}$	4.007	4.006	4.007	4.004	4.003	4.004
se($\widehat{\Lambda_0(25)}$)	0.4033	0.3935	0.3939	0.3945	0.3866	0 .3870
e.se($\widehat{\Lambda_0(25)}$)	0.3826	0.3825	0.3824	0.3793	0.3790	0.3789
$\widehat{\Lambda_0(48)}$	5.550	5.549	5.551	5.552	5.551	5.552
se($\widehat{\Lambda_0(48)}$)	0.5508	0.5421	0.5427	0.5469	0.5368	0 .5374
e.se($\widehat{\Lambda_0(48)}$)	0.5269	0.5266	0.5264	0.5239	0.5235	0.5233
$\widehat{\beta}$	1.5043	1.5045	1.5041	1.5042	1.5045	1.5041
se($\widehat{\beta}$)	0.1229	0.1202	0.1205	0.1229	0.1202	0.1204
e.se($\widehat{\beta}$)	0.1214	0.1202	0.1217	0.1214	0.1218	0.1217
$\widehat{v}$	0.4902	0.4788	0.4743	0.4899	0.4782	0.4736
se($\widehat{v}$)	0.1068	0.1068	0.1153	0.1067	0.1067	0.1151
e.se($\widehat{v}$)	0.1027	0.1210	0.1390	0.1028	0.1209	0.1390

Table 2.2: The sample average of estimates and estimated standard errors based on asymptotic theory, and the empirical standard errors, in setting (b) $\gamma = 0.8$, $\delta = 0.5$, $\beta = 1.5$, $v = 0.2$, $\Lambda_0(25) = 4$, $\Lambda_0(48) = 5.543$

	I	II-1	II-2	III	IV-1	IV-2
$\widehat{\Lambda_0(25)}$	4.002	4.002	4.003	3.993	3.993	3.994
se($\widehat{\Lambda_0(25)}$)	0.329	0.327	0.328	0.318	0.317	0.318
e.se($\widehat{\Lambda_0(25)}$)	0.3318	0.3322	0.3316	0.3081	0.3085	0.3079
$\widehat{\Lambda_0(48)}$	5.542	5.542	5.543	5.536	5.535	5.537
se($\widehat{\Lambda_0(48)}$)	0.445	0.444	0.444	0.440	0.439	0.440
e.se($\widehat{\Lambda_0(48)}$)	0.439	0.439	0.438	0.4257	0.426	0.425
$\widehat{\beta}$	1.5025	1.5026	1.5020	1.5025	1.5025	1.5020
se($\widehat{\beta}$)	0.0968	0.0963	0.0966	0.0968	0.0963	0.0966
e.se($\widehat{\beta}$)	0.0989	0.0989	0.0990	0.0989	0.0990	0.0991
$\widehat{v}$	0.198	0.195	0.198	0.198	0.196	0.198
se($\widehat{v}$)	0.0544	0.0517	0.0611	0.0545	0.0518	0.0612
e.se($\widehat{v}$)	0.0538	0.0606	0.0653	0.0538	0.0605	0.0652

Table 2.3: The sample average of estimates and estimated standard errors based on asymptotic theory, and the empirical standard errors, in setting (g) $\gamma = 0.8$, $\delta = 1.0$, $\beta = 0.375$, $v = 0.5$, $\Lambda_0(25) = 20$, $\Lambda_0(48) = 38.4$

	I	II-1	II-2	III	IV-1	IV-2
$\widehat{\Lambda_0(25)}$	19.88	19.88	19.88	19.90	19.90	19.90
se($\widehat{\Lambda_0(25)}$)	1.56	1.55	1.55	1.53	1.52	1.52
e.se($\widehat{\Lambda_0(25)}$)	1.65	1.65	1.65	1.58	1.58	1.58
$\widehat{\Lambda_0(48)}$	38.26	38.26	38.26	38.27	38.26	38.27
se($\widehat{\Lambda_0(48)}$)	2.96	2.94	2.94	2.94	2.92	2.92
e.se($\widehat{\Lambda_0(48)}$)	3.10	3.10	3.10	3.11	3.11	3.11
$\widehat{\beta}$	0.373	0.373	0.373	0.373	0.373	0.373
se($\widehat{\beta}$)	0.0939	0.0931	0.0931	0.0939	0.0931	0.0931
e.se($\widehat{\beta}$)	0.0816	0.0816	0.0816	0.0816	0.0816	0.0817
$\widehat{v}$	0.489	0.484	0.485	0.489	0.485	0.485
se($\widehat{v}$)	0.0763	0.0839	0.0840	0.0763	0.0840	0.0841
e.se($\widehat{v}$)	0.0700	0.104	0.104	0.0699	0.104	0.105

Table 2.4: The sample average of estimates and estimated standard errors based on asymptotic theory, and the empirical standard errors, in setting (h) $\gamma = 0.8$, $\delta = 1.0$, $\beta = 0.375$, $v = 0.2$, $\Lambda_0(25) = 20$, $\Lambda_0(48) = 38.4$

	I	II-1	II-2	III	IV-1	IV-2
$\widehat{\Lambda_0(25)}$	20.21	20.21	20.21	20.27	20.27	20.27
$se(\widehat{\Lambda_0(25)})$	1.10	1.08	1.08	1.05	1.04	1.04
$e.se(\widehat{\Lambda_0(25)})$	1.03	1.03	1.03	0.951	0.951	0.950
$\widehat{\Lambda_0(48)}$	38.91	38.91	38.91	38.89	38.89	38.89
$se(\widehat{\Lambda_0(48)})$	2.04	2.02	2.02	2.01	2.00	2.00
$e.se(\widehat{\Lambda_0(48)})$	1.80	1.80	1.80	1.86	1.86	1.86
$\widehat{\beta}$	0.3701	0.3702	0.3701	0.3701	0.3702	0.3701
$se(\widehat{\beta})$	0.0631	0.0620	0.0620	0.0631	0.0620	0.0620
$e.se(\widehat{\beta})$	0.0728	0.0729	0.0729	0.0728	0.0729	0.0729
$\widehat{v}$	0.200	0.197	0.197	0.200	0.197	0.197
$se(\widehat{v})$	0.0352	0.0359	0.0361	0.0352	0.0359	0.0361
$e.se(\widehat{v})$	0.0373	0.0396	0.0395	0.0372	0.0394	0.0393

Table 2.5: The actual coverage probabilities(x100) of $(1-\alpha)$ confidence intervals based on 100 samples for setting (a)

	I		II-1		II-2		III		IV-1		IV-2	
$100(1-\alpha)$	95	90	95	90	95	90	95	90	95	90	95	90
$\Lambda_0(25)$	94	91	94	92	94	92	94	91	95	92	95	92
$\Lambda_0(48)$	95	91	95	90	95	90	95	91	94	91	94	91
β	93	89	94	88	94	88	93	89	94	88	94	88
v	95	91	87	83	88	87	95	91	87	83	88	87
$\log v$	94	92	88	84	91	88	94	92	88	84	89	88

Table 2.6: The actual coverage probabilities(x100) of $(1-\alpha)$ confidence intervals based on 100 samples for setting (g)

	I		II-1		II-2		III		IV-1		IV-2	
$100(1-\alpha)$	95	90	95	90	95	90	95	90	95	90	95	90
$\Lambda_0(25)$	92	91	92	90	92	90	92	91	92	91	92	91
$\Lambda_0(48)$	92	89	92	89	92	89	92	89	92	89	92	89
β	96	92	97	92	97	92	96	91	97	92	97	92
v	95	92	84	77	84	77	96	92	85	77	85	77
$\log v$	98	92	87	82	88	82	98	92	87	82	88	82

Table 2.7: Maximum likelihood estimates in the mixed Poisson process models $(\hat{\rho}, \hat{\beta}, \hat{v})$ and using generalized estimating equations $(\tilde{\rho}, \tilde{\beta}, \tilde{v})$

ρ

interval	(0,5.5]	(5.5,10.5]	(10.5,15.5]	(15.5,20.5]	(20.5,25.5]	(25.5,30.5]	(30.5, 40.5]	(40.5,53]
$\hat{\rho}$	0.134	0.0722	0.0895	0.0657	0.142	0.0798	0.118	0.0430
se($\hat{\rho}$)	0.060	0.034	0.042	0.032	0.065	0.040	0.054	0.024
$\tilde{\rho}$	0.134	0.0725	0.0900	0.0661	0.143	0.0795	0.117	0.0429
se($\tilde{\rho}$)	0.059	0.038	0.054	0.037	0.073	0.042	0.061	0.029

β

parameter	β_1	β_2	β_3
$\hat{\beta}$	-1.220	0.379	-0.00998
se($\hat{\beta}$)	0.376	0.104	0.129
$\tilde{\beta}$	-1.211	0.376	-0.00931
se($\tilde{\beta}$)	0.320	0.0872	0.105

v

	ML	GEE
estimate	2.37	1.85
se	0.50	0.40

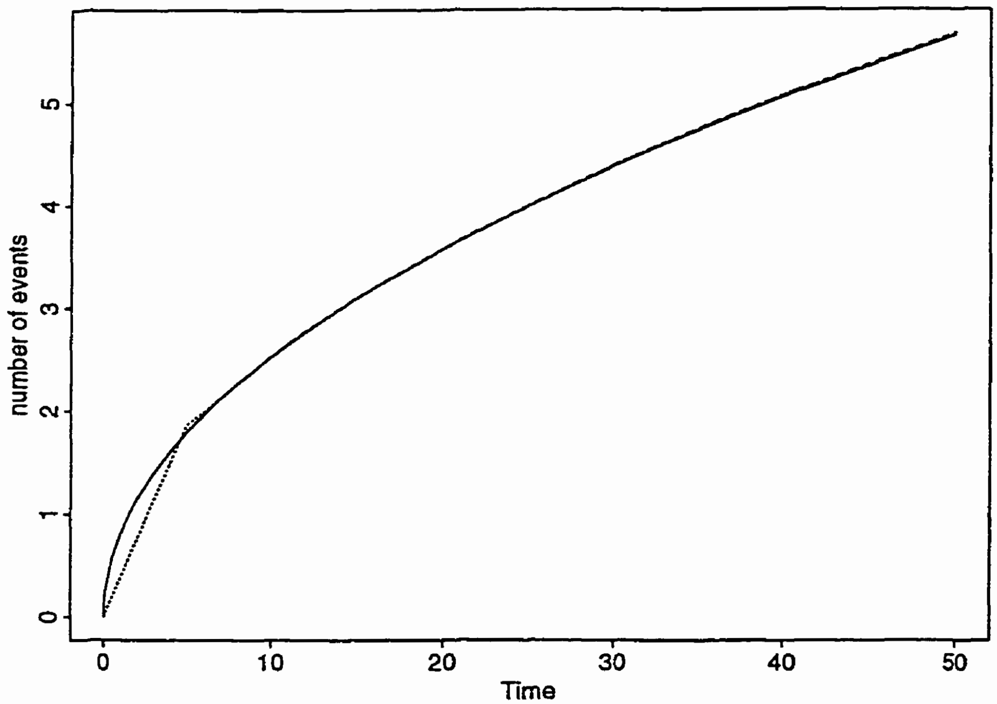

Figure 2.1: The true baseline mean function and its estimates for setting (a). The solid curve is the true baseline mean function; the dotted curve is the estimated mean function based on Approach I; the dashed curve, which is indistinguishable from the solid curve, is the estimated mean function based on Approach III.

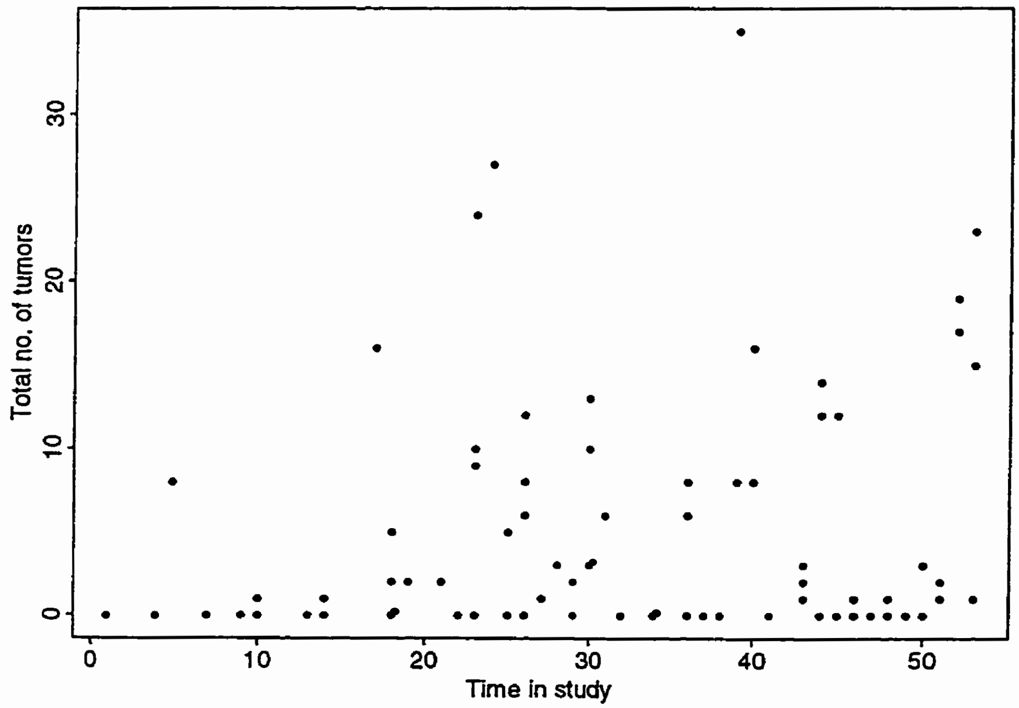

Figure 2.2: Scatter plot of the total number of tumors for each patient against time in study

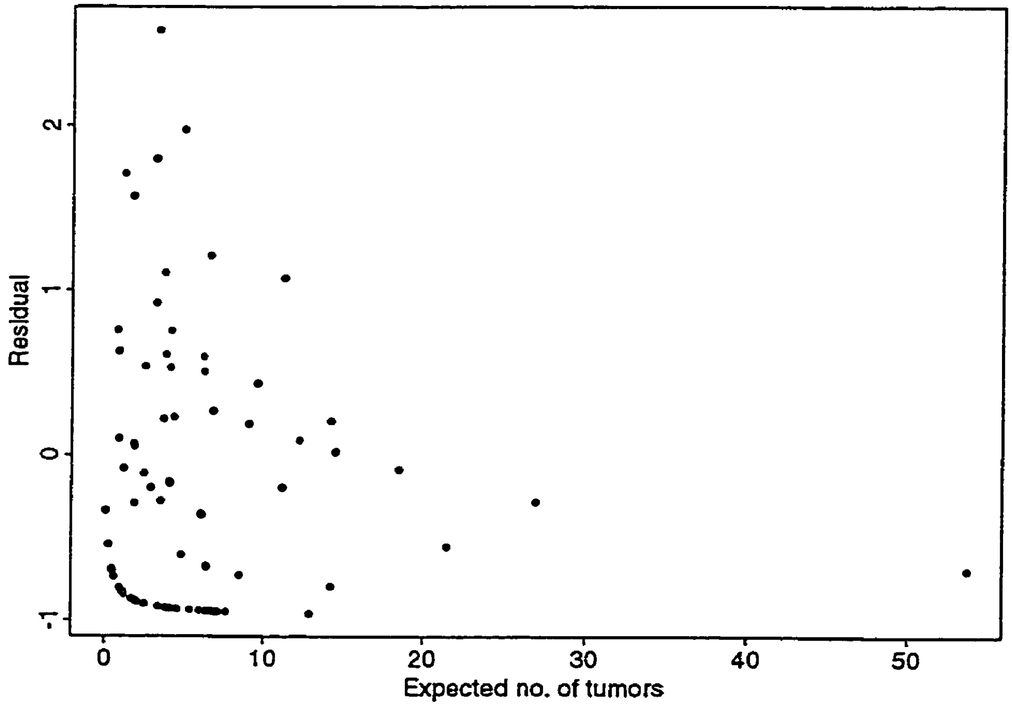

Figure 2.3: Plot of the standardized residual e_i against the expected number of tumors present, based on a mixed Poisson model

2.7 Appendix

Tables 2.8 to 2.11 present the sample average of estimates and the estimated standard errors, and the empirical standard errors for simulation in settings (c), (d), (e) and (f).

Tables 2.12 to 2.17 present the empirical coverage of 90% and 95% confidence intervals based on simulations in settings (b), (c), (d), (e), (f) and (h).

Table 2.8: The sample average of estimates and estimated standard errors based on asymptotic theory, and the empirical standard errors, in setting (c) $\gamma = 0.8$, $\delta = 1.0$, $\beta = 1.5$, $v = 0.5$, $\Lambda_0(25) = 20$, $\Lambda_0(48) = 38.4$

	I	II-1	II-2	III	IV-1	IV-2
$\widehat{\Lambda_0(25)}$	19.773	19.773	19.774	19.725	19.725	19.726
se($\widehat{\Lambda_0(25)}$)	1.568	1.536	1.536	1.544	1.517	1.518
e.se($\widehat{\Lambda_0(25)}$)	1.601	1.603	1.603	1.584	1.585	1.585
$\widehat{\Lambda_0(48)}$	37.890	37.890	37.892	37.899	37.899	37.901
se($\widehat{\Lambda_0(48)}$)	2.976	2.928	2.928	2.967	2.917	2.917
e.se($\widehat{\Lambda_0(48)}$)	2.979	2.981	2.981	2.994	2.9967	2.996
$\widehat{\beta}$	1.4905	1.4905	1.4904	1.4905	1.4905	1.4904
se($\widehat{\beta}$)	0.0970	0.0950	0.0950	0.0969	0.0950	0.0950
e.se($\widehat{\beta}$)	0.0985	0.0984	0.0984	0.0984	0.0984	0.0984
$\widehat{v}$	0.480	0.462	0.462	0.480	0.462	0.462
se($\widehat{v}$)	0.0782	0.0818	0.0824	0.0782	0.0818	0.0824
e.se($\widehat{v}$)	0.0667	0.0806	0.0828	0.0668	0.0807	0.0829

Table 2.9: The sample average of estimates and estimated standard errors based on asymptotic theory, and the empirical standard errors, in setting (d) $\gamma = 0.8$, $\delta = 1.0$, $\beta = 1.5$, $v = 0.2$, $\Lambda_0(25) = 20$, $\Lambda_0(48) = 38.4$

	I	II-1	II-2	III	IV-1	IV-2
$\widehat{\Lambda_0(25)}$	19.952	19.950	19.952	20.002	20.000	20.001
se($\widehat{\Lambda_0(25)}$)	1.101	1.110	1.110	1.075	1.081	1.081
e.se($\widehat{\Lambda_0(25)}$)	0.951	0.951	0.951	0.947	0.947	0.947
$\widehat{\Lambda_0(48)}$	38.43	38.43	38.43	38.40	38.39	38.40
se($\widehat{\Lambda_0(48)}$)	2.077	2.082	2.084	2.062	2.070	2.071
e.se($\widehat{\Lambda_0(48)}$)	1.868	1.868	1.868	1.832	1.832	1.833
$\widehat{\beta}$	1.500	1.500	1.500	1.500	1.500	1.500
se($\widehat{\beta}$)	0.0668	0.0660	0.0661	0.0668	0.0660	0.0661
$\widehat{v}$	0.191	0.190	0.191	0.191	0.190	0.191
se($\widehat{v}$)	0.0353	0.0362	0.0381	0.0353	0.0363	0.0382
e.se($\widehat{v}$)	0.0365	0.0357	0.0371	0.0366	0.0360	0.0373

Table 2.10: The sample average of estimates and estimated standard errors based on asymptotic theory, and the empirical standard errors, in setting (e) $\gamma = 0.8$, $\delta = 0.5$, $\beta = 0.375$, $v = 0.5$, $\Lambda_0(25) = 4$, $\Lambda_0(48) = 5.543$

	I	II-1	II-2	III	IV-1	IV-2
$\widehat{\Lambda_0(25)}$	3.940	3.940	3.940	3.936	3.936	3.936
se($\widehat{\Lambda_0(25)}$)	0.365	0.363	0.363	0.348	0.345	0.345
e.se($\widehat{\Lambda_0(25)}$)	0.376	0.376	0.376	0.352	0.352	0.352
$\widehat{\Lambda_0(48)}$	5.450	5.450	5.451	5.452	5.452	5.453
se($\widehat{\Lambda_0(48)}$)	0.489	0.484	0.484	0.482	0.478	0.478
e.se($\widehat{\Lambda_0(48)}$)	0.468	0.468	0.468	0.476	0.476	0.476
$\widehat{\beta}$	0.3654	0.3654	0.3653	0.3655	0.3655	0.3655
se($\widehat{\beta}$)	0.108	0.107	0.107	0.108	0.107	0.107
e.se($\widehat{\beta}$)	0.0982	0.0981	0.982	0.0979	0.0979	0.0979
$\widehat{v}$	0.470	0.458	0.459	0.471	0.459	0.460
se($\widehat{v}$)	0.108	0.109	0.110	0.108	0.109	0.110
e.se($\widehat{v}$)	0.109	0.127	0.130	0.108	0.127	0.129

Table 2.11: The sample average of estimates and estimated standard errors based on asymptotic theory, and the empirical standard errors, in setting (f) $\gamma = 0.8$, $\delta = 0.5$, $\beta = 0.375$, $v = 0.2$, $\Lambda_0(25) = 4$, $\Lambda_0(48) = 5.543$

	I	II-1	II-2	III	IV-1	IV-2
$\widehat{\Lambda_0(25)}$	3.993	3.993	3.993	3.968	3.968	3.968
se($\widehat{\Lambda_0(25)}$)	0.293	0.290	0.290	0.270	0.268	0.268
e.se($\widehat{\Lambda_0(25)}$)	0.336	0.336	0.336	0.297	0.297	0.297
$\widehat{\Lambda_0(48)}$	5.490	5.490	5.490	5.493	5.493	5.494
se($\widehat{\Lambda_0(48)}$)	0.382	0.378	0.378	0.373	0.372	0.372
e.se($\widehat{\Lambda_0(48)}$)	0.416	0.416	0.416	0.411	0.411	0.411
$\widehat{\beta}$	0.3661	0.3661	0.3661	0.3662	0.3662	0.3662
se($\widehat{\beta}$)	0.0827	0.0824	0.0825	0.0827	0.0824	0.0825
e.se($\widehat{\beta}$)	0.0883	0.0882	0.0881	0.0883	0.0882	0.0881
$\widehat{v}$	0.187	0.186	0.185	0.187	0.186	0.185
se($\widehat{v}$)	0.0613	0.0606	0.0616	0.0613	0.0605	0.0614
e.se($\widehat{v}$)	0.0640	0.0659	0.0656	0.0638	0.0657	0.0653

Table 2.12: The actual coverage probabilities(x100) of $(1-\alpha)$ confidence intervals based on 100 samples for setting (b)

	I		II-1		II-2		III		IV-1		IV-2	
$100(1-\alpha)$	95	90	95	90	95	90	95	90	95	90	95	90
$\Lambda_0(25)$	97	89	96	87	96	87	96	94	96	92	96	92
$\Lambda_0(48)$	96	92	96	88	96	88	96	94	96	91	96	91
β	96	89	95	91	95	91	96	89	95	91	95	91
v	95	91	90	85	96	89	95	91	90	84	96	90

Table 2.13: The actual coverage probabilities(x100) of $(1-\alpha)$ confidence intervals based on 100 samples for setting (c)

	I		II-1		II-2		III		IV-1		IV-2	
$100(1-\alpha)$	95	90	95	90	95	90	95	90	95	90	95	90
$\Lambda_0(25)$	92	86	91	84	91	84	92	82	92	81	92	81
$\Lambda_0(48)$	93	83	94	83	94	83	93	83	93	82	93	82
β	95	85	93	83	93	83	95	85	93	83	93	83
v	94	93	89	80	89	80	94	93	88	80	89	78

Table 2.14: The actual coverage probabilities(x100) of $(1-\alpha)$ confidence intervals based on 100 samples for setting (d)

	I		II-1		II-2		III		IV-1		IV-2	
$100(1-\alpha)$	95	90	95	90	95	90	95	90	95	90	95	90
$\Lambda_0(25)$	98	95	98	96	98	96	97	95	97	94	97	94
$\Lambda_0(48)$	97	94	96	93	96	93	97	95	97	94	97	94
β	90	89	91	88	91	88	90	89	91	88	91	88
v	90	86	85	83	87	84	90	82	85	83	87	84

Table 2.15: The actual coverage probabilities(x100) of $(1-\alpha)$ confidence intervals based on 100 samples for setting (e)

	I		II-1		II-2		III		IV-1		IV-2	
$100(1-\alpha)$	95	90	95	90	95	90	95	90	95	90	95	90
$\Lambda_0(25)$	94	88	93	89	93	89	94	89	94	89	94	89
$\Lambda_0(48)$	94	92	93	92	93	92	93	88	93	88	93	88
β	97	94	97	92	97	92	97	94	97	93	97	93
v	90	87	81	75	84	77	89	87	83	74	85	76

Table 2.16: The actual coverage probabilities(x100) of $(1-\alpha)$ confidence intervals based on 100 samples for setting (f)

	I		II-1		II-2		III		IV-1		IV-2	
$100(1-\alpha)$	95	90	95	90	95	90	95	90	95	90	95	90
$\Lambda_0(25)$	93	83	91	83	91	83	92	89	91	90	91	90
$\Lambda_0(48)$	94	87	92	85	92	85	91	87	91	87	91	87
$\boldsymbol{\beta}$	94	87	94	86	94	86	94	88	94	86	94	86
v	90	87	87	82	87	82	90	87	86	82	85	83

Table 2.17: The actual coverage probabilities(x100) of $(1-\alpha)$ confidence intervals based on 100 samples for setting (h)

	I		II-1		II-2		III		IV-1		IV-2	
$100(1-\alpha)$	95	90	95	90	95	90	95	90	95	90	95	90
$\Lambda_0(25)$	95	93	95	92	95	92	96	91	97	91	97	91
$\Lambda_0(48)$	97	92	97	93	97	93	97	90	97	90	97	90
$\boldsymbol{\beta}$	92	87	91	87	91	87	92	86	91	87	91	87
v	94	90	92	85	93	87	94	91	92	86	93	86

Chapter 3

Analysis of Current Status Data and Doubly-censored Data

3.1 Introduction

In many studies, it is of interest to estimate the distribution function of the time between two successive events, termed the initiating event and subsequent event. Let I and J represent the occurrence times of the two events, respectively, then $T = J - I$ is the time between the two events. We assume in this chapter that I and T are independent. If both I and J are observed directly, estimation of distribution of T can be readily obtained. However, it is not an easy task when one or both events are not directly observed. The analysis of current status data and doubly-censored data fall into this category of data collection. Current status data arise when the time of the initiating event is observed directly, but for the subsequent event the only information is whether or not it has occurred by a single monitoring time B. This type of data is often collected in a survey or census, where respondents are asked about their age and whether a certain event (usually an important one, such as first marriage) has occurred.

A more complicated data structure occurs when both events are not observed directly; instead, we observe only whether either or both events have occurred at the single monitoring time B. This is called doubly-censored current status data, an extreme case of doubly-censored data. For example, in the Partners' Study described by Jewell *et al.* (1994), the initiating event is infection of an individual with HIV, and the subsequent event is the subsequent infection of a sexual partner of this individual. If the only information we know is that the first infection time lies in a time interval (A, B) and whether the second infection has occurred at time B, the data are doubly censored current status data. General doubly-censored data refer to the situation where either or both events are not observed directly, but rather the initiating event time is only known to lie in an interval, and the subsequent event time is interval-censored or right censored. The data structures can also include covariates in these problems.

Current status data and doubly censored data have attracted considerable attention. Diamond and Mcdonald (1992) have discussed advantages and disadvantages of current status data and have reviewed the fitting of parametric proportional hazard models, parametric accelerated life models and semiparametric proportional hazard models to such data. Since the larger number of parameters in the semiparametric approach may create difficulties in model fitting, they suggest using a spline form for a suitable transformation of the baseline cumulative hazard function. Sun and Kalbfleisch (1993) discussed statistical methods for current status data coming from point processes. Jewell and van der Laan (1996) have reviewed both parametric and semiparametric methods for extensions of current status data, such as doubly censored current status data and current status information on more complicated stochastic processes. They have pointed out that nonparametric maximum likelihood estimation (NPMLE) of the distribution function G of $T = J - I$ in the doubly censored current status data case can be obtained by viewing the model as a nonparametric mixture estimation problem; however, the NPMLE of G may be inconsistent. De Gruttola and Lagakos (1989) have proposed methods for analyzing general doubly censored data in the absence of covariates. Kim *et al.* (1993) have generalized the results of De Gruttola and Lagakos (1989) to incorporate covariates. However, they assume that I and T are discrete. Sun (1995) proposed a self-consistency algorithm to obtain the non- parametric estimation of a distribution function with truncated and doubly censored data.

In this chapter, we intend to develop weakly parametric methods for current status data and doubly censored data by assuming a piecewise constant form for the hazard function of T. In more detail, we assume that there is a pre-specified sequence of constants $a_1 = 0 < a_2 < \ldots < a_r < a_{r+1} = \infty$ that divide the time axis into r intervals, and assume that the hazard function $h_0(t)$ of T is constant within interval $A_k = (a_k, a_{k+1}]$: let $h_0(t) = \rho_k$ for $t \in A_k$. The cumulative hazard function is $H_0(t) = \sum_{k=1}^{r} \rho_k u_k(t)$, where

$$u_k(t) = \begin{cases} 0 & \text{if } t \leq a_k \\ t - a_k & \text{if } a_k < t \leq a_{k+1} \\ a_{k+1} - a_k & \text{otherwise} \end{cases} \tag{3.1}$$

The survival function of T is $S_0(t) = P(T > t) = exp(-H_0(t))$, and the cumulative distribution function of T is $G_0(t) = 1 - S_0(t) = 1 - exp(-H_0(t))$.

The advantage of using a piecewise constant form for the hazard function is that it avoids the difficulty in estimating standard errors of estimates in purely non-parametric models and at the same time provides a more flexible model than most parametric models.

Covariate effects can be assessed by using regression models. Different regression models are considered for current status data and doubly censored data in this chapter.

3.2 Standard Current Status Data

3.2.1 Estimate of a CDF in the absence of covariates

First we describe the data structure of standard current status data in more detail. Following Jewell and van der Laan (1996), suppose that there are n independent subjects (or n pairs of subjects) in our study. For subject i, at recruitment time (or interview) time B_i, we know that the initiating event has already occurred and this event time I_i is observed. The subject is checked at monitoring time B_i to see whether the subsequent event J_i has occurred or not. Thus we observe whether the time between the two events $T_i = J_i - I_i$ is $\leq B_i - I_i$ or $> B_i - I_i$. We assume that the initiating event time I_i and the time between the two events T_i are independent, and the monitoring time B_i is also independent of I_i and T_i. The data consists of observations (C_i, δ_i), $i = 1, \ldots, n$ where $C_i = B_i - I_i$, and $\delta_i = 1$ if $J_i \leq B_i$ and is zero otherwise. Therefore for $i = 1, \ldots, n$, $1 - \delta_i$ is a Bernoulli variate with success probability $\mu_i = Pr(1 - \delta_i = 1) = P_i = S_0(C_i)$, where $S_0(t)$ is the survival function of T. The likelihood function based on the data (conditional on the observed values of the C_i's) is $L = \prod_{i=1}^{n}(1-P_i)^{\delta_i} P_i^{1-\delta_i}$. Suppose m is the number of distinct C_i's, and the m distinct values of C_i's are $C_{(1)} < C_{(2)} < \ldots < C_{(m)}$. Let $s_{(i)}$ denote the subset of indices corresponding to $C_{(i)}$; that is, Subject j is in set $s_{(i)}$, if $C_j = C_{(i)}$. Let the size of $s_{(i)}$ be k_i, let $P_{(i)} = S_0(C_{(i)})$, and let $Y_i = \sum_{j \in s_i} \delta_i$. Then the likelihood can be simplified to $L = \prod_{j=1}^{m}(1 - P_{(j)})^{Y_j} P_{(j)}^{k_j - Y_j}$. The sufficient statistics are $(Y_1, \cdots, Y_m)$. Therefore for the model with piecewise constant intensity the number of a_k's can be no more than m, and the location of a_k's is restricted by the location of $C_{(i)}$'s. More specifically, in order to be able to estimate all ρ_k (by that, I mean the set of the likelihood equations has unique solution for ρ_k's), a_k's must satisfy the following necessary conditions: (1) The last interval $A_r = (a_r, \infty)$ contains at least one C_i, or equivalently, $a_r < C_{(m)}$. (2) There must be at least one $C_{(j)}$ in the union of any two consecutive intervals, i.e., (a_k, a_{k+2}), for $k = 1, \ldots, r-1$, assuming $a_{r+1} = \infty$. Proof of Condition (1). If the last interval A_r does not contain any of the C_j's, i.e.,

$a_r \geq C_{(m)}$. So $u_r(C_{(k)}) = 0$, for any $k = 1, \ldots, m$. So

$$P_{(k)} = exp\{-\sum_{j=1}^{r} \rho_j u_j(C_{(k)})\} = exp\{-\sum_{j=1}^{r-1} \rho_j u_j(C_{(k)})\},$$

for $k = 1, \ldots, m$. So the likelihood function does not involve ρ_r. ρ_r cannot be estimated from this model.

Proof of Condition (2). If $k = r-1$, then by Condition (1), there must be at least one $C_{(j)}$ in the last interval (a_r, ∞), so there must be at least one $C_{(j)}$ in the interval (a_{r-1}, ∞). Suppose that $k < r-1$, and there are $C_{(1)}, \ldots, C_{(m_1)}$ in interval $(a_1, a_k]$, no C_j's in interval (a_k, a_{k+2}), and there are $C_{(m_1)}, \ldots, C_{(m)}$ in interval $[a_{k+2}, \infty)$, where $m_1 < m$. We have

$$\begin{aligned} P_{(j)} &= exp\{-\sum_{l=1}^{k-1} \rho_l u_l(C_{(j)})\}, \quad j = 1, \ldots, m_1, \qquad (3.2) \\ P_{(j)} &= exp\{-\sum_{l=1}^{k-1} \rho_l(a_{l+1} - a_l) - \rho_k(a_{k+1} - a_k) - \rho_{k+1}(a_{k+2} - a_{k+1})\} \\ &\quad exp\{-\sum_{l=k+2}^{r} \rho_l u_l(C_{(j)})\}, \quad j = m_1 + 1, \ldots, m. \end{aligned}$$

So ρ_k and ρ_{k+1} cannot be estimated separately. To see it, we compute the likelihood equations for ρ_k and ρ_{k+1}. The likelihood functions are

$$\frac{\partial logL}{\partial \rho_i} = \sum_{j=1}^{m} [k_j - Y_j/(1 - P_{(j)})] \frac{1}{P_{(j)}} \frac{\partial P_{(j)}}{\partial \rho_i}, \quad i = 1, \ldots, r. \qquad (3.3)$$

Since

$$\frac{\partial P_{(j)}}{\partial \rho_k} = 0, \quad j = 1, \ldots, m_1,$$

$$\frac{\partial P_{(j)}}{\partial \rho_k} = -(a_{k+1} - a_k)P_{(j)}, \quad j = m_1 + 1, \ldots, m,$$

the likelihood equation for ρ_k is

$$\frac{\partial logL}{\partial \rho_k} = -(a_{k+1} - a_k) \sum_{j=m_1}^{m} [k_j - Y_j/(1 - P_{(j)})] = 0. \tag{3.4}$$

Similarly, we can show the likelihood equation for ρ_{k+1} is

$$\frac{\partial logL}{\partial \rho_{k+1}} = -(a_{k+2} - a_{k+1}) \sum_{j=m_1}^{m} [k_j - Y_j/(1 - P_{(j)})] = 0. \tag{3.5}$$

Both equations are the same. So there cannot be an unique solution to the set of the likelihood equations unless other assumptions are made.

If m is large, we can choose r between 5 to 10 and let each A_k interval contain some $C_{(i)}$'s.

The log-likelihood function is

$$\log L = \sum_{i=1}^{n} (\delta_i \log(1 - P_i) + (1 - \delta_i) \log P_i),$$

where $P_i = exp\{-\sum_{k=1}^{r} \rho_k u_k(C_i)\}$. The score function is

$$\begin{aligned} \frac{\partial \log L}{\partial \rho_k} &= \sum_{i=1}^{n} \frac{1 - \delta_i - P_i}{(1 - P_i)P_i} \frac{\partial P_i}{\partial \rho_k} \\ &= \sum_{i=1}^{n} \frac{-u_k(C_i)(1 - \delta_i - P_i)}{1 - P_i} \\ &= \sum_{i=1}^{n} (\delta_i u_k(C_i)/(1 - P_i) - u_k(C_i)), \end{aligned} \tag{3.6}$$

$k = 1, \ldots, r.$

The second derivatives of the log-likelihood function are

$$\frac{\partial^2 \log L}{\partial \rho_k \partial \rho_j} = -\sum_{i=1}^{n} \frac{\delta_i P_i u_k(C_i) u_j(C_i)}{(1-P_i)^2}, \tag{3.7}$$

$$k, j = 1, \ldots, r. \tag{3.8}$$

Therefore, the (k, j) entry of the Fisher information matrix is

$$E(-\frac{\partial^2 \log L}{\partial \rho_k \partial \rho_j}) = \sum_{i=1}^{n} \frac{P_i u_k(C_i) u_j(C_i)}{1-P_i}. \tag{3.9}$$

The maximum likelihood estimates of ρ_k's can be found using the Newton-Raphson algorithm. Since we have $\log P_i = \log(S_0(C_i)) = -\sum_{k=1}^{r} \rho_k u_k(C_i)$, the model is a generalized linear model with logarithmic link function and Bernoulli distribution. The parameters ρ_k can therefore also be estimated by many statistical programs, such as SAS or Splus. Let $\hat{\rho}_k$ be the estimates. The survival function $S_0(t)$ can be estimated by $\hat{S}_0(t) = \exp(-\sum_{k=1}^{r} \hat{\rho}_k u_k(t))$.

3.2.2 Regression models

If we have recorded a vector of baseline covariates $z_i = (z_{i1}, \cdots, z_{ip})$ for Subject i, the relationship between the time T and the covariates can be studied. Carstensen (1996) has discussed several regression models with piecewise constant hazards in the context of interval censored lifetime data. Since standard current status data is a special case of interval censored lifetime data, the models discussed there can be applied to standard current status data. We briefly describe the fitting of these models for current status data.

The additive excess risk model assumes a hazard function $h(t|z_i) = h_0(t) + z_i'\beta$, with the constraint $h_0(t) + z_i \geq 0$. The likelihood function is

$$L = \prod_{i=1}^{n} (1 - P_i)^{\delta_i} P_i^{1-\delta_i}, \tag{3.10}$$

where $P_i = S(C_i|z_i)$ and

$$S(t|z_i) = \exp\left\{ -\sum_{k=1}^{r} \rho_k u_k(t) - \sum_{j=1}^{p} \beta_j (z_{ij} t) \right\}.$$

This model is also a generalized linear model with logarithmic link function and Bernoulli distribution.

The proportional hazard model (multiplicative relative risk model) has a hazard function $h(t|z_i) = h_0(t) exp(z_i'\beta)$. The likelihood function has the same form as (3.10), but with $S(t|z_i) = \exp\{-H_0(t)\exp(z_i'\beta)\}$. The parameter estimates can be obtained by an iteration procedure given by Carstensen (1996), or by Newton's method.

The fully parametric and semi parametric accelerated failure time models (AFT) for current status data or interval censored lifetime data are discussed by some authors such as Diamond and Mcdonald (1992), Rabinowitz *et al.* (1995), and Jewell and van der Laan (1996). We outline some ideas on an accelerated failure time model with a piecewise constant density in a finite interval for the random errors. We assume that

$$\log T_i = z_i'\beta + \epsilon_i, \tag{3.11}$$

where ϵ_i's are independent random errors whose common density function is symmetric about zero and piecewise constant in interval $[a_1, a_{r+1}]$. In detail, we assume that r is an odd number, $(a_1, \cdots, a_{r+1})$ are pre-specified constants, such that $a_k = -a_{r+2-k}$. Let $A_k = (a_k, a_{k+1}]$. The density function of ϵ is given by

$$f(t) = \begin{cases} f_k & \text{if } t \in A_k \text{ or } A_{r+1-k}, k = 1, \cdots, (r+1), \\ 0 & \text{if } t > a_{r+1} \text{ or } t < a_1 \end{cases}$$

So $f_k = f_{r+2-k}$, $k = 1, \cdots, r+1$. f_k must satisfy the constraint $\sum_{k=1}^{r} f_k(a_{k+1} - a_k) = 1$. The number of distinct parameters is $(r-1)/2$. The cumulative distribution function of ϵ_i is $F(t) = \sum_{k=1}^{r} f_k u_k(t)$. Thus

$$Pr(\delta_i = 1) = Pr(\epsilon_i \leq \log C_i - z_i'\beta)$$
$$= F(\log C_i - z_i'\beta) = \sum_{k=1}^{r} f_k u_k(\log C_i - z_i'\beta).$$

The likelihood function can be obtained as before and an optimization algorithm can be used to get the maximum likelihood estimates.

3.2.3 Simulated examples

To illustrate the proposed methods, we applied the methods to several simulated datasets. Since for the standard current status data, the initiating event time I is observed, we need only to consider the induction time $T = J - I$, the observation time for T, which is $C = B - I$, and possibly a covariate vector z. This is equivalent to setting $I = 0$, $T = J$ and $C = B$. The way of generating data was very similar to the one in Shiboski (1998). That is, (1) a single covariate z was generated from a Bernoulli distribution with $p = 0.5$; (2) observation times C were generated from a uniform [0,1] distribution; (3) failure times T were generated from a Weibull proportional hazard model, such that the hazard function is of the form,

$$h(t|z) = \alpha/\gamma(t/\gamma)^{\alpha-1}\exp(z\beta).$$

The value of the indicator variable δ was determined according to the observed T and C. Both the value of β and the value of γ were set at one. Three values were used for α: $0.5, 1.1, 2$. A dataset of size 100 was generated for each value of α.

We focus on estimation of the regression coefficient β. The proportional hazard models with a piecewise constant baseline hazard function and a Weibull baseline hazard function were fitted to each dataset. The downhill simplex method due to Nelder and Mead (see, Press *et al.*, 1990) was used to get the maximum likelihood estimates of the parameters. The number of pieces was set to be five or ten, with cutoff points being $0, 0.2, 0.4, 0.6, 0.8$ for the five-piece model and $0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ for the ten-piece model.

The estimates of β, the estimated standard errors (based on the Fisher information matrix) and the log-likelihood values at the maximum likelihood estimates are given in Table 3.1. Both piecewise constant models give very similar values of $\hat{\beta}$. These values are also reasonably close to the estimates given by the Weibull model. It suggests that estimates of β are robust to the number of pieces used in the piecewise constant models. However, the piecewise constant models are less efficient than the Weibull model (the true parametric model) for estimating β.

Table 3.1: Estimation of regression parameter β based on proportional hazard models with piecewise constant hazard and Weibull hazard.

		$\alpha = 0.5$	$\alpha = 1.1$	$\alpha = 2.0$
Weibull	$\hat{\beta} - \beta$	-0.157	0.0168	-0.238
	$se(\hat{\beta})$	0.275	0.320	0.341
	-logL	55.66	50.86	45.12
PC (5 pieces)	$\hat{\beta} - \beta$	-0.00206	0.0358	-0.216
	$se(\hat{\beta})$	0.305	0.342	0.358
	-logL	55.19	50.05	44.73
PC (10 pieces)	$\hat{\beta} - \beta$	-0.0208	0.0951	-0.187
	$se(\hat{\beta})$	0.306	0.358	0.367
	-logL	54.34	48.27	42.68

3.3 Doubly Censored Current Status Data

As we have pointed out in Section 3.1, doubly censored current status data arise when the initiating event time I is known only to be in an interval (X_L, X_R), and it is also observed whether or not the subsequent event has occurred at time B; that is, we observe $\delta = 1$ if $J \leq B$, and $\delta = 0$ if $J > B$. Let $T = J - I$. We make the following assumptions: (1) I and T are independent; (2) the observation times and censoring times are independent of I and T; (3) the distribution of I is known. Let $W(t)$ and $w(t)$ be the cumulative distribution function (CDF) and density function for I respectively. Our goal is to estimate the survival function (call it $S_0(t)$) of the time between the two event times, based on an independent sample $(X_{Li}, X_{Ri}, B_i, \delta_i)$, for $i = 1, \ldots, n$. If a vector of baseline covariates is measured for each subject, we can also study the effect of covariates on the distribution of T.

3.3.1 A piecewise constant hazard model

Suppose we have n independent observations $(X_{Li}, X_{Ri}, B_i, \delta_i)$, $i = 1, \ldots, n$. We assume that $X_{Ri} \leq B_i$, for $i = 1, \ldots, n$. Under the assumptions (1) to (3), the likelihood is $L = \prod_{i=1}^{n}(1-P_i)^{\delta_i} P_i^{1-\delta_i}$, where $P_i = Pr(J_i > B_i | I_i \in [X_{Li}, X_{Ri}])$. Now for simplicity, drop the index i and just use P. It is easy to see that

$$P = \int_{X_L}^{X_R} S_0(B-x) w^*(x) dx, \tag{3.12}$$

where $w^*(x) = w(x) / \int_{X_L}^{X_R} w(t)dt$ is the conditional density of I in interval $[X_L, X_R]$. We assume $w^*(x)$ is known. If we further assume that w^* is uniform on the given interval $[X_L, X_R]$, P can be simplified as

$$\begin{aligned} P &= \int_{X_L}^{X_R} S_0(B-x)dx/(X_R - X_L) \\ &= \int_{B-X_R}^{B-X_L} S_0(x)dx/(X_R - X_L). \end{aligned} \tag{3.13}$$

In particular, if $X_R = B$,

$$P = \int_0^{X_R - X_L} S_0(x)dx/(X_R - X_L).$$

Under the assumption of a piecewise constant hazard for T, we have $S_0(t) = \exp(-H_0(t)) = \exp(-\sum_{k=1}^{r} \rho_k u_k(t))$, where $H_0(t)$, ρ_k and $u_k(t)$ are defined in Section 3.1. For $a_j \leq t \leq a_{j+1}$, let $P(t|a_j) = Pr(T > t | T > a_j) = S_0(t)/S_0(a_j) = exp(-\rho_j(t - a_j))$. We compute $\int_0^c S_0(x)dx$ for a given constant $c > 0$. There exists an integer $j(c)$, such that $a_{j(c)} < c \leq a_{j(c)+1}$, and $1 \leq j(c) \leq r$. It can be shown that

$$\begin{aligned}\int_0^c S_0(x)dx &= \sum_{j=1}^{j(c)-1} \rho_j^{-1}[S_0(a_j) - S_0(a_{j+1})] + \rho_{j(c)}^{-1}[S_0(a_{j(c)}) - S_0(c)] \\ &= \sum_{j=1}^{j(c)-1} \rho_j^{-1} S_0(a_j)[1 - P(a_{j+1}|a_j)] + \rho_{j(c)}^{-1} S_0(a_{j(c)})[1 - P(c|a_{j(c)})] \\ &= \sum_{j=1}^{r} \rho_j^{-1} S_0(a_j)[1 - exp(-\rho_j u_j(c))]. \qquad (3.14)\end{aligned}$$

3.3.2 Regression Models

Suppose for each subject, we also observe z_i, a vector of covariates that is related to the distribution of T. The likelihood is still given by $L = \prod_{i=1}^{n}(1 - P_i)^{\delta_i} P_i^{1-\delta_i}$, where $P_i = Pr(J_i > X_{Ri}|I_i \in [X_{Li}, X_{Ri}], z_i)$, if the observation time $B_i = X_{Ri}$. The additive excess risk model assumes a hazard function $h(t|z_i) = h_0(t) + z_i'\beta$. So we have

$$S(t|z_i) = exp(-H_0(t) - tz_i'\beta) = exp(-\sum_{k=1}^{r}(\rho_k + z_i'\beta)u_k(t)) = S_0(t)exp(-tz_i'\beta),$$

and

$$\begin{aligned}\int_0^c S(t|z_i)dt &= \sum_{j=1}^{j(c)-1} (\rho_j + z_i'\beta)^{-1} S(a_j|z_i)[1 - exp(-(\rho_j + z_i'\beta)(a_{j+1} - a_j))] \\ &+ (\rho_{j(c)} + z_i'\beta)^{-1} S(a_{j(c)}|z_i)[1 - exp(-(\rho_{j(c)} + z_i'\beta)(c - a_{j(c)}))],\end{aligned}$$

where $j(c)$ is an integer, such that $a_{j(c)} < c \leq a_{j(c)+1}$, and $1 \leq j(c) \leq r$. Thus $P_i = \int_0^{X_{Ri}-X_{Li}} S(t|z_i)dt/(X_{Ri} - X_{Li})$.

A derivative-free optimization method, such as the downhill simplex method due to Nelder and Mead (see, Press *et al.*, 1990), can be applied to obtain the maximum likelihood estimates of (ρ, β). In fact, this is the method we used in our simulations and examples in Chapter 3.

The proportional hazard model assumes a hazard function for T_i is

$$h(t|z_i) = h_0(t)exp(z_i'\beta). \tag{3.15}$$

Let $d_i = exp(z_i'\beta)$, then $S(c|z_i) = S_0(c)^{d_i} = exp(-H_0(c)d_i)$. Suppose c is in interval $(a_{j(c)}, a_{j(c)+1}]$,

$$\begin{aligned}
IS(c, z_i) &= \int_0^c S(t|z_i)dt \qquad (3.16)\\
&= \sum_{j=1}^{j(c)-1} \rho_j^{-1} d_i^{-1}[S(a_j|z_i) - S(a_{j+1}|z_i)] + \rho_{j(c)}^{-1} d_i^{-1}[S(a_{j(c)}|z_i) - S(c|z_i)]\\
&= \sum_{j=1}^{j(c)-1} \rho_j^{-1} d_i^{-1} exp(-\sum_{k=1}^{j-1} \rho_k d_i |A_k|)[1 - exp(-\rho_j d_i |A_j|)]\\
&+ \rho_{j(c)}^{-1} d_i^{-1} exp(-\sum_{k=1}^{j(c)-1} \rho_k d_i (a_{k+1} - a_k))[1 - exp(-\rho_{j(c)} d_i (c - a_{j(c)})]\\
&= \sum_{k=1}^{r} \rho_k^{-1} d_i^{-1} S(a_k|z_i)[1 - \exp(-\rho_k d_i u_k(c))].
\end{aligned}$$

$IS(c, 0) = \int_0^c S_0(t)dt$. So $P_i = IS(X_{Ri} - X_{Li}, z_i)/(X_{Ri} - X_{Li})$ for the proportional hazard model.

The score functions for either regression model are:

$$\begin{aligned}
\frac{\partial logL}{\partial \rho_k} &= \sum_{i=1}^{n} \frac{(1 - \delta_i - P_i)}{P_i(1 - P_i)} \frac{\partial P_i}{\partial \rho_k}, \qquad (3.17)\\
k &= 1, \ldots, r \qquad (3.18)
\end{aligned}$$

$$\frac{\partial logL}{\partial \beta_j} = \sum_{i=1}^{n} \frac{(1-\delta_i - P_i)}{P_i(1-P_i)} \frac{\partial P_i}{\partial \beta_j}, \tag{3.19}$$

$$j = 1, \ldots, r. \tag{3.20}$$

The Fisher Information matrix is

$$I_F = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}$$

where the (k, j) element of I_{11} is

$$E(-\frac{\partial^2 logL}{\partial \rho_k \partial \rho_j}) = \sum_{i=1}^{n} \frac{\partial P_i}{\partial \rho_k} \frac{\partial P_i}{\partial \rho_j}, \tag{3.21}$$

the (k, j) element of I_{12} is

$$E(-\frac{\partial^2 logL}{\partial \rho_k \partial \beta_j}) = \sum_{i=1}^{n} \frac{\partial P_i}{\partial \rho_k} \frac{\partial P_i}{\partial \beta_j}, \tag{3.22}$$

the (k, j) element of I_{22} is

$$E(-\frac{\partial^2 logL}{\partial \beta_k \partial \beta_j}) = \sum_{i=1}^{n} \frac{\partial P_i}{\partial \beta_k} \frac{\partial P_i}{\partial \beta_j}, \tag{3.23}$$

and $I_{21} = I_{12}^t$. The derivatives of P_i's with respect to ρ_k's and β_j's for the proportional hazard model are given in an appendix in Section 3.8.

The maximum likelihood estimates of (ρ, β) can be obtained by Fisher's scoring method. An alternative is to use a derivative-free optimization method.

3.4 General Doubly Censored Data

We now consider data with general double censoring and left truncation. First we introduce some notation. Suppose there are n independent subjects in our study and that Subject i is enrolled at time E_i. One of the conditions to be included in the study is that the subsequent event time J_i has not occurred at the time of enrollment E_i; that is, $J_i > E_i$. The initiating event time I_i is observed to be in interval $[X_{Li}, X_{Ri}]$. The subsequent event time J_i is observed to be in interval $[J_{Li}, J_{Ri}]$, where $J_{Li} > E_i$. Let $T_i = J_i - I_i$, and assume I_i and T_i are independent and continuous. Let $W_i(x)$, $w_i(x)$ be the cumulative distribution function and density function of I_i, respectively, for $x \in (0, \infty)$. Let $G_i(t)$, $S_i(t)$, $g_i(t)$ and $h_i(t)$ be the cumulative distribution function, survival function, density function and hazard function of T_i.

Assuming that I_i's do not contain information on the distribution of T_i, the likelihood function conditional on $I_i \in [X_{Li}, X_{Ri}]$ and $J_i > E_i$ is

$$L = \prod_{i=1}^{n} \frac{\int_{X_{Li}}^{X_{Ri}} w_i^*(x)[S_i(J_{Li} - x) - S_i(J_{Ri} - x)]dx}{\int_{X_{Li}}^{X_{Ri}} w_i^*(x) S_i(E_i - x)dx}, \tag{3.24}$$

where $w_i^*(x) = w_i(x) / \int_{X_{Li}}^{X_{Ri}} w_i(x)dx$, $X_{Li} \le x \le X_{Ri}$, is the density of I_i conditional on $I_i \in [X_{Li}, X_{Ri}]$. We assume $S_i(x) = 0$ if $x \le 0$.

If $E_i \le X_{Li}$, there is no truncation, and the denominator of the likelihood function is 1. Also, the case of right-censored J_i's can be treated as a special case of our general formulation with $J_{Ri} = \infty$. Suppose $g_i(t)$ is the density function of

T_i. The likelihood function can be written as

$$L = \prod_{i=1}^{n} \frac{\int_{X_{Li}}^{X_{Ri}} w_i^*(x) g_i(J_{Li} - x)^{\delta_i} S_i(J_{Li} - x)^{1-\delta_i} dx}{\int_{X_{Li}}^{X_{Ri}} w_i^*(x) S_i(E_i - x) dx}, \tag{3.25}$$

where δ_i is the censoring indicator for J_i. $\delta_i = 1$ if J_{Li} is the observed value for J_i; $\delta_i = 0$ if J_{Li} is the observed censoring time.

In the rest of this section, we assume that $w_i^*(x)$ is known.

3.4.1 A piecewise constant hazard model

We assume that $w_i^*(x)$ is uniform over interval $[X_{Li}, X_{Ri}]$. The baseline hazard function for T is a piecewise constant function $h_0(t) = \rho_k$, if $t \in A_k$, where $A_k = (a_k, a_{k+1}]$ and $0 = a_1 < a_2 < \ldots < a_r < a_{r+1} = \infty$ is a sequence of pre-fixed constants. We also observe a vector of covariates z_i for each subject. We assume the covariate effect on the hazard function of T is multiplicative, that is, a proportional hazard model is used:

$$h(t|z_i) = h_0(t) exp(z_i'\beta). \tag{3.26}$$

Let G_0, S_0 and g_0 be the baseline cumulative distribution function, survival function, and density function. Recall that $S_0(t) = exp(-H_0(t)) = exp(-\sum_{k=1}^{r} \rho_k u_k(t))$, $G_0(t) = 1 - S_0(t)$, $g_0(t) = h_0(t) exp(-H_0(t))$, where $u_k(t)$ is defined in formula (3.1).

Now we compute the log-likelihood based on (3.25). It can be written as

$$logL = \sum_{i=1}^{n} [log(B_{1i}) - log(B_{2i})], \tag{3.27}$$

where

$$
\begin{aligned}
B_{1i} &= \frac{1}{X_{Ri}-X_{Li}}\int_{X_{Li}}^{X_{Ri}} h(J_{Li}-x|z_i)^{\delta_i} S(J_{Li}-x|z_i)dx \\
&= \frac{1}{X_{Ri}-X_{Li}}\int_{J_{Li}-X_{Ri}}^{J_{Li}-X_{Li}} h(x|z_i)^{\delta_i} S(x|z_i)dx \\
&= \frac{1}{X_{Ri}-X_{Li}}[S(J_{Li}-X_{Ri}|z_i)-S(J_{Li}-X_{Li}|z_i)]^{\delta_i} \\
&\quad [IS(J_{Li}-X_{Li},z_i)-IS(J_{Li}-X_{Ri},z_i)]^{1-\delta_i}, \qquad (3.28) \\
B_{2i} &= \frac{1}{X_{Ri}-X_{Li}}\int_{X_{Li}}^{X_{Ri}} S(E_i-x|z_i)dx \\
&= \frac{1}{X_{Ri}-X_{Li}}\int_{E_i-X_{Ri}}^{E_i-X_{Li}} S(x|z_i)dx \\
&= \frac{1}{X_{Ri}-X_{Li}}[IS(E_i-X_{Li},z_i)-IS(E_i-X_{Ri},z_i)], \qquad (3.29)
\end{aligned}
$$

where $IS(c,z)$ is defined in (3.16). So

$$
\begin{aligned}
logL &= \sum_{i=1}^{n}\{\delta_i log[S(J_{Li}-X_{Ri}|z_i)-S(J_{Li}-X_{Li}|z_i)] \\
&\quad +(1-\delta_i)log[IS(J_{Li}-X_{Li},z_i)-IS(J_{Li}-X_{Ri},z_i)] \\
&\quad -log[IS(E_i-X_{Li},z_i)-IS(E_i-X_{Ri},z_i)]\}. \qquad (3.30)
\end{aligned}
$$

3.4.2 Computation

The maximum likelihood estimates (MLEs) of ρ_k's can be obtained by maximizing the log likelihood function. This can be achieved by a derivative-free optimization method, or an optimization method using derivatives. The method we used is the downhill simplex method due to Nelder and Mead (see Press *et al.*, 1990). After MLEs are obtained, the standard errors of the estimates can be derived from the observed information matrix, that is, the minus second derivative matrix of the log likelihood function with respect to the parameters, evaluated at the MLEs. We give the first and second derivatives here.

To simplify our notation, let

$$C_{i1} = IS(J_{Li} - X_{Li}, z_i), \tag{3.31}$$

$$C_{i2} = IS(J_{Li} - X_{Ri}, z_i), \tag{3.32}$$

$$C_{i3} = IS(E_i - X_{Li}, z_i), \tag{3.33}$$

$$C_{i4} = IS(E_i - X_{Ri}, z_i), \tag{3.34}$$

$$C_{i5} = S(J_{Li} - X_{Ri}|z_i), \tag{3.35}$$

$$C_{i6} = S(J_{Li} - X_{Li}|z_i). \tag{3.36}$$

Then

$$logL = \sum_{i=1}^{n}\{\delta_i log(C_{i5} - C_{i6}) + (1-\delta_i)log(C_{i1} - C_{i2}) - log(C_{i3} - C_{i4})\}. \tag{3.37}$$

The first derivatives of $logL$ with respect to ρ and β are:

$$\begin{aligned} \frac{\partial logL}{\partial \rho_k} &= \sum_{i=1}^{n}\{\frac{\delta_i}{C_{i5} - C_{i6}}(\frac{\partial C_{i5}}{\partial \rho_k} - \frac{\partial C_{i6}}{\partial \rho_k}) + \frac{(1-\delta_i)}{C_{i1} - C_{i2}}(\frac{\partial C_{i1}}{\partial \rho_k} - \frac{\partial C_{i2}}{\partial \rho_k}) \\ &\quad - \frac{1}{C_{i3} - C_{i4}}(\frac{\partial C_{i3}}{\partial \rho_k} - \frac{\partial C_{i4}}{\partial \rho_k})\}, \end{aligned} \tag{3.38}$$

$$\begin{aligned} \frac{\partial logL}{\partial \beta_j} &= \sum_{i=1}^{n}\{\frac{\delta_i}{C_{i5} - C_{i6}}(\frac{\partial C_{i5}}{\partial \beta_j} - \frac{\partial C_{i6}}{\partial \beta_j}) + \frac{(1-\delta_i)}{C_{i1} - C_{i2}}(\frac{\partial C_{i1}}{\partial \beta_j} - \frac{\partial C_{i2}}{\partial \beta_j}) \\ &\quad - \frac{1}{C_{i3} - C_{i4}}(\frac{\partial C_{i3}}{\partial \beta_j} - \frac{\partial C_{i4}}{\partial \beta_j})\}. \end{aligned} \tag{3.39}$$

The second derivatives of $logL$ with respect to ρ and β are:

$$\frac{\partial^2 logL}{\partial \rho_k^2} = \sum_{i=1}^{n}\{\frac{\delta_i}{C_{i5}-C_{i6}}(\frac{\partial^2 C_{i5}}{\partial \rho_k^2}-\frac{\partial^2 C_{i6}}{\partial \rho_k^2}) - \frac{\delta_i}{(C_{i5}-C_{i6})^2}(\frac{\partial C_{i5}}{\partial \rho_k}-\frac{\partial C_{i6}}{\partial \rho_k})^2$$
$$+ \frac{(1-\delta_i)}{C_{i1}-C_{i2}}(\frac{\partial^2 C_{i1}}{\partial \rho_k^2}-\frac{\partial^2 C_{i2}}{\partial \rho_k^2}) - \frac{(1-\delta_i)}{(C_{i1}-C_{i2})^2}(\frac{\partial C_{i1}}{\partial \rho_k}-\frac{\partial C_{i2}}{\partial \rho_k})^2$$
$$- \frac{1}{C_{i3}-C_{i4}}(\frac{\partial^2 C_{i3}}{\partial \rho_k^2}-\frac{\partial^2 C_{i4}}{\partial \rho_k^2}) + \frac{1}{(C_{i3}-C_{i4})^2}(\frac{\partial C_{i3}}{\partial \rho_k}-\frac{\partial C_{i4}}{\partial \rho_k})^2\} \quad (3.40)$$

$$\frac{\partial^2 logL}{\partial \rho_k \partial \rho_j} = \sum_{i=1}^{n}\{\frac{\delta_i}{C_{i5}-C_{i6}}(\frac{\partial^2 C_{i5}}{\partial \rho_k \partial \rho_j}-\frac{\partial^2 C_{i6}}{\partial \rho_k \partial \rho_j})$$
$$-\frac{\delta_i}{(C_{i5}-C_{i6})^2}(\frac{\partial C_{i5}}{\partial \rho_k}-\frac{\partial C_{i6}}{\partial \rho_k})(\frac{\partial C_{i5}}{\partial \rho_j}-\frac{\partial C_{i6}}{\partial \rho_j})$$
$$+\frac{1-\delta_i}{C_{i1}-C_{i2}}(\frac{\partial^2 C_{i1}}{\partial \rho_k \partial \rho_j}-\frac{\partial^2 C_{i2}}{\partial \rho_k \partial \rho_j})$$
$$-\frac{(1-\delta_i)}{(C_{i1}-C_{i2})^2}(\frac{\partial C_{i1}}{\partial \rho_k}-\frac{\partial C_{i2}}{\partial \rho_k})(\frac{\partial C_{i1}}{\partial \rho_j}-\frac{\partial C_{i2}}{\partial \rho_j})$$
$$-\frac{1}{(C_{i3}-C_{i4})}(\frac{\partial^2 C_{i3}}{\partial \rho_k \partial \rho_j}-\frac{\partial^2 C_{i4}}{\partial \rho_k \partial \rho_j})$$
$$+\frac{1}{(C_{i3}-C_{i4})^2}(\frac{\partial C_{i3}}{\partial \rho_k}-\frac{\partial C_{i4}}{\partial \rho_k})(\frac{\partial C_{i3}}{\partial \rho_j}-\frac{\partial C_{i4}}{\partial \rho_j})\}, \quad (3.41)$$

for $k \neq j$,

$$\frac{\partial^2 logL}{\partial \rho_k \partial \beta_j} = \sum_{i=1}^{n}\{\frac{\delta_i}{C_{i5}-C_{i6}}(\frac{\partial^2 C_{i5}}{\partial \rho_k \partial \beta_j}-\frac{\partial^2 C_{i6}}{\partial \rho_k \partial \beta_j})$$
$$-\frac{\delta_i}{(C_{i5}-C_{i6})^2}(\frac{\partial C_{i5}}{\partial \rho_k}-\frac{\partial C_{i6}}{\partial \rho_k})(\frac{\partial C_{i5}}{\partial \beta_j}-\frac{\partial C_{i6}}{\partial \beta_j})$$
$$+\frac{1-\delta_i}{C_{i1}-C_{i2}}(\frac{\partial^2 C_{i1}}{\partial \rho_k \partial \beta_j}-\frac{\partial^2 C_{i2}}{\partial \rho_k \partial \beta_j})$$
$$-\frac{(1-\delta_i)}{(C_{i1}-C_{i2})^2}(\frac{\partial C_{i1}}{\partial \rho_k}-\frac{\partial C_{i2}}{\partial \rho_k})(\frac{\partial C_{i1}}{\partial \beta_j}-\frac{\partial C_{i2}}{\partial \beta_j})$$
$$-\frac{1}{(C_{i3}-C_{i4})}(\frac{\partial^2 C_{i3}}{\partial \rho_k \partial \beta_j}-\frac{\partial^2 C_{i4}}{\partial \rho_k \partial \beta_j})$$
$$+\frac{1}{(C_{i3}-C_{i4})^2}(\frac{\partial C_{i3}}{\partial \rho_k}-\frac{\partial C_{i4}}{\partial \rho_k})(\frac{\partial C_{i3}}{\partial \beta_j}-\frac{\partial C_{i4}}{\partial \beta_j})\}, \quad (3.42)$$

$$\begin{aligned}\frac{\partial^2 logL}{\partial\beta_k\partial\beta_j} &= \sum_{i=1}^{n}\{\frac{\delta_i}{C_{i5}-C_{i6}}(\frac{\partial^2 C_{i5}}{\partial\beta_k\partial\beta_j}-\frac{\partial^2 C_{i6}}{\partial\beta_k\partial\beta_j})\\ &-\frac{\delta_i}{(C_{i5}-C_{i6})^2}(\frac{\partial C_{i5}}{\partial\beta_k}-\frac{\partial C_{i6}}{\partial\beta_k})(\frac{\partial C_{i5}}{\partial\beta_j}-\frac{\partial C_{i6}}{\partial\beta_j})\\ &+\frac{1-\delta_i}{C_{i1}-C_{i2}}(\frac{\partial^2 C_{i1}}{\partial\beta_k\partial\beta_j}-\frac{\partial^2 C_{i2}}{\partial\beta_k\partial\beta_j})\\ &-\frac{(1-\delta_i)}{(C_{i1}-C_{i2})^2}(\frac{\partial C_{i1}}{\partial\beta_k}-\frac{\partial C_{i2}}{\partial\beta_k})(\frac{\partial C_{i1}}{\partial\beta_j}-\frac{\partial C_{i2}}{\partial\beta_j})\\ &-\frac{1}{(C_{i3}-C_{i4})}(\frac{\partial^2 C_{i3}}{\partial\beta_k\partial\beta_j}-\frac{\partial^2 C_{i4}}{\partial\beta_k\partial\beta_j})\\ &+\frac{1}{(C_{i3}-C_{i4})^2}(\frac{\partial C_{i3}}{\partial\beta_k}-\frac{\partial C_{i4}}{\partial\beta_k})(\frac{\partial C_{i3}}{\partial\beta_j}-\frac{\partial C_{i4}}{\partial\beta_j})\}. \qquad (3.43)\end{aligned}$$

3.4.3 A Weibull regression model

Suppose for example that T_i follows a Weibull distribution with a hazard function $h_i(t;z_i) = h_0(t)\exp(z_i'\beta)$, and $h_0(t) = \alpha\gamma^{-\alpha}t^{\alpha-1}$. The baseline survival function is $S_0(t) = \exp\{-(t/\gamma)^\alpha\}$. We still assume that $w_i^*(x)$ is uniform over interval $[X_{Li}, X_{Ri}]$. For the same data structure as in 3.4.1, the log likelihood based on observed data is

$$\begin{aligned}logL &= \sum_{i=1}^{n}\{\delta_i log[S(J_{Li}-X_{Ri}|z_i)-S(J_{Li}-X_{Li}|z_i)]\\ &+(1-\delta_i)log[IS(J_{Li}-X_{Li},z_i)-IS(J_{Li}-X_{Ri},z_i)]\\ &-log[IS(E_i-X_{Li},z_i)-IS(E_i-X_{Ri},z_i)]\}. \qquad (3.44)\end{aligned}$$

Now $S(t|z_i) = exp(-(t/\gamma)^\alpha d_i)$, where $d_i = exp(z_i'\beta)$.

$$IS(t,z_i) = \int_0^t \exp\{-(x/\gamma)^\alpha d_i\}dx$$

$$= \frac{\gamma}{\alpha}(1/d_i)^{1/\alpha} \int_0^{(t/\gamma)^\alpha d_i} \exp(-y) y^{1/\alpha - 1} dy$$
$$= (\gamma/\alpha)(1/d_i)^{1/\alpha}\Gamma(1/\alpha)P(1/\alpha, (t/\gamma)^\alpha d_i), \tag{3.45}$$

where

$$P(a,x) = (1/\Gamma(a)) \int_0^x y^{a-1} e^{-y} dy$$

is the incomplete Gamma function. Notice that

$$\Gamma(a+1,x)P(a+1,x) = a\Gamma(a,x)P(a,x) - exp(-x)x^a.$$

The formulas for derivatives of $S(t|z_i)$ and $IS(t, z_i)$ are in an appendix (Section 3.8). These are used to obtain variance estimates for the parameters. An alternative would be to use numerical derivatives or the alternative covariance matrix estimate

$$\hat{V}(\hat{\rho}, \hat{\beta}) = \left(\sum_{i=1}^{n} (\frac{\partial log L_i}{\partial \theta})(\frac{\partial log L_i}{\partial \theta^T}) \right)_{\hat{\theta}}^{-1},$$

based on the fact that

$$E\left[(\frac{\partial log L_i}{\partial \theta})(\frac{\partial log L_i}{\partial \theta})^T \right] = E\left\{ -\frac{\partial log L_i}{\partial \theta \partial \theta^T} \right\}.$$

3.5 A Simulation Study

We conduct a simulation study to assess the performance of the piecewise constant hazard models and to examine the effect of the lengths of the intervals for the first event (I_i) on estimation.

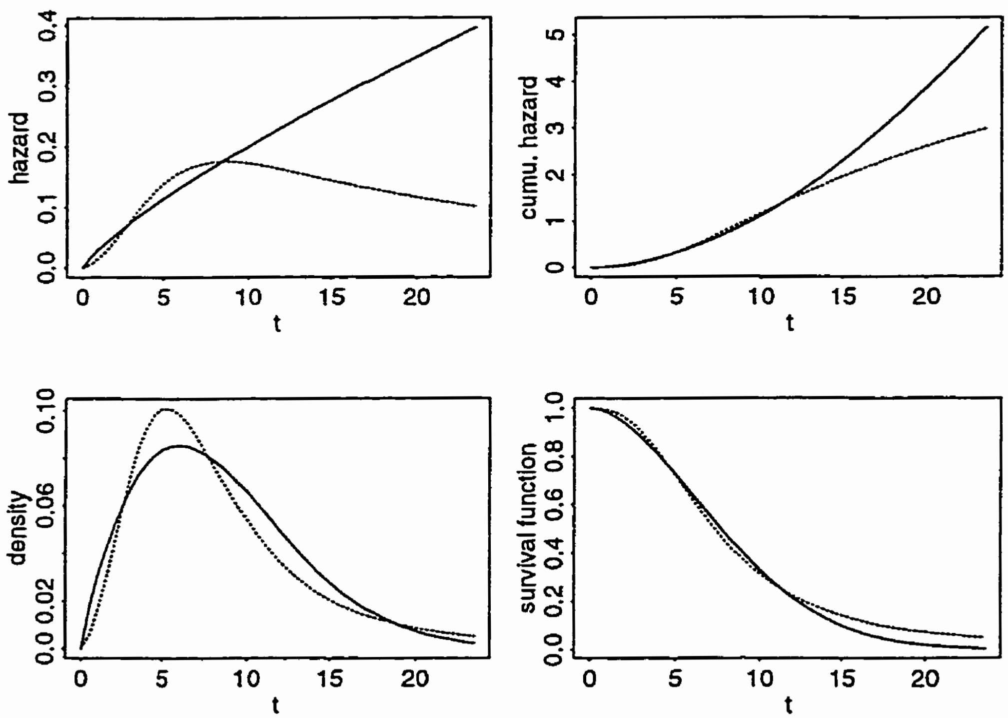

Figure 3.1: Hazard functions, cumulative hazard functions, density functions and survival functions of Weibull and log-logistic distributions used in the simulation. Solid lines–Weibull; dotted lines–log logistic

We mimic an AIDS cohort study in our simulation set-up. Suppose subjects are under periodic examination (screening) to check whether the first event has occurred or not during time interval $(0, \tau]$. The periodic examination times are $\omega, 2\omega, \ldots, m\omega$, where $m = \tau/\omega$. Once a subject has experienced the first event between two consecutive screening times, this subject is followed up from the screening time right after the first event time, until the end of the study period B ($B > \tau$) to check when the second event occurs. We only consider subjects with first events occurred during interval $(0,\tau]$. The sample size is chosen to be $n = 100$. The data are generated in the following steps. (1) We generate the first event times I_i, $i = 1,\ldots,n$ from the uniform distribution on interval $[0,\tau]$. (2) We generate T_i, $i = 1,\ldots,n$, from a continuous distribution with CDF $G(t)$. (3) Let $J_i = I_i + T_i$ be the second event time for subject i. There are three forms of outcomes. The first one is that both I_i and J_i $(I_i < J_i)$ are known only to be in interval $(k\omega,(k+1)\omega]$, for some $k = 1,\ldots,m-1$; the second is that I_i is only known to be in interval $(k\omega,(k+1)\omega]$ (for some k), and $J_i > (k+1)\omega$ is observed exactly; the third is that I_i is known only to be in interval $(k\omega,(k+1)\omega]$ (for some k), and J_i is known only to be greater than B. Therefore the observed data vector for subject i is $(X_{Li}, X_{Ri}, J_{Li}, \delta_{1i}, \delta_{2i})$, where $X_{Li} = k\omega$, $X_{Ri} = (k+1)\omega$, $J_{Li} = max(X_{Ri}, min(J_i, B))$, $\delta_{1i} = 1$ if $J_i > X_{Ri}$; $\delta_{1i} = 0$ otherwise. $\delta_{2i} = 1$ if $J_i \leq B$; $\delta_{2i} = 0$ otherwise. So, (1) for the first outcome, $X_{Li} = k\omega$, $X_{Ri} = J_{Li} = (k+1)\omega, \delta_{1i} = 0$, $\delta_{2i} = 1$; (2) for the second outcome, $X_{Li} = k\omega$, $X_{Ri} = (k+1)\omega$, $J_{Li} = J_i$, $\delta_{1i} = \delta_{2i} = 1$; (3) for the third outcome, $X_{Li} = k\omega$, $X_{Ri} = (k+1)\omega$, $J_{Li} = B$, $\delta_{1i} = 1$, $\delta_{2i} = 0$.

We set $\tau = 6$. Two distributions are considered for T_i's, the time between the two events, a Weibull distribution with CDF $G(t) = 1 - exp(-(t/\gamma)^\alpha)$, where $\gamma = 9.434$, $\alpha = 1.8$; a log-logistic distribution with CDF $G(t) = 1 - 1/(1 + (\gamma t)^\alpha)$, where $\gamma = 0.1368365$, $\alpha = 2.515052$. The parameters in the log-logistic distribution were chosen to make it have the same first and third quartiles as the given Weibull distribution. The plots of the hazard functions, cumulative hazard functions, density functions and survival functions of the given Weibull and log-

logistic distributions are given in Figure 3.1.

Two values of ω are used: 1 or 6. The end of follow-up time is $B = \tau/2 + G^{-1}(0.9)$, which gives 17.99 for the Weibull distribution, and 20.507 for the log-logistic distribution. It results in four cases: (1) T follows a Weibull distribution, $\omega = 1$, $B = 17.99$; (2) T follows a Weibull distribution, $\omega = 6$, $B = 17.99$; (3) T follows a log-logistic distribution, $\omega = 1$, $B = 20.507$; (4) T follows a log-logistic distribution, $\omega = 6$, $B = 20.507$. For each case, 500 data sets were generated.

The likelihood based on the observed data is similar to (3.25) but without the truncation:

$$\begin{aligned} L &= \prod_{i=1}^{n} \int_{X_{Li}}^{X_{Ri}} g(J_{Li} - x)^{\delta_{1i}\delta_{2i}} G(X_{Ri} - x)^{\delta_{2i}(1-\delta_{1i})} \\ &\quad S(J_{Li} - x)^{\delta_{1i}(1-\delta_{2i})} w_i^*(x) dx \\ &= \prod_{i=1}^{n} \int_{X_{Li}}^{X_{Ri}} h(J_{Li} - x)^{\delta_{1i}\delta_{2i}} S(J_{Li} - x)^{\delta_{1i}} \\ &\quad [1 - S(X_{Ri} - x)]^{\delta_{2i}(1-\delta_{1i})} w_i^*(x) dx, \end{aligned} \tag{3.46}$$

where $g(t)$, $h(t)$, $G(t)$ and $S(t)$ are density function, hazard function, cumulative distribution function and survival function of the T_i's, respectively; $w_i^*(x)$ is the density of I_i conditional on $I_i \in [X_{Li}, X_{Ri}]$. Here since I_i follows a Uniform distribution, $w_i^*(x) = 1/(X_{Ri} - X_{Li})$, for $x \in [X_{Li}, X_{Ri}]$.

The likelihood function can also be expressed in the following form:

$$\begin{aligned} L &= \prod_{i=1}^{n} \frac{1}{X_{Ri} - X_{Li}} [S(J_{Li} - X_{Ri}) - S(J_{Li} - X_{Li})]^{\delta_{1i}\delta_{2i}} \\ &\quad [X_{Ri} - X_{Li} - IS(X_{Ri} - X_{Li})]^{\delta_{2i}(1-\delta_{1i})} \end{aligned}$$

$$[IS(J_{Li} - X_{Li}) - IS(J_{Li} - X_{Ri})]^{\delta_{1i}(1-\delta_{2i})}, \tag{3.47}$$

where $IS(t) = \int_0^t S(x)dx$.

The corresponding log-likelihood function (ignoring a constant term) is

$$\begin{aligned} \log L &= \sum_{i=1}^{n} \delta_{1i}\delta_{2i} \log\left[S(J_{Li} - X_{Ri}) - S(J_{Li} - X_{Li})\right] \\ &\quad + \sum_{i=1}^{n} (1-\delta_{1i})\delta_{2i} \log\left[X_{Ri} - X_{Li} - IS(X_{Ri} - X_{Li})\right] \\ &\quad + \sum_{i=1}^{n} \delta_{1i}(1-\delta_{2i}) \log\left[IS(J_{Li} - X_{Li}) - IS(J_{Li} - X_{Ri})\right], \\ &= \sum_{i=1}^{n} (\delta_{1i}\delta_{2i} \log(C_{i5} - C_{i6}) + (1-\delta_{1i})\delta_{2i} \log(X_{Ri} - X_{Li} - C_{i7}) \\ &\quad + \delta_{1i}(1-\delta_{2i}) \log(C_{i1} - C_{i2})), \end{aligned} \tag{3.48}$$

where $C_{i1}, C_{i2}, C_{i5}, C_{i6}$ are defined in equations (3.31), to (3.36); C_{i7} is defined as

$$C_{i7} = IS(X_{Ri} - X_{Li}, z_i).$$

In general, suppose $\boldsymbol{\theta}$ is the p-dimensional parameter in the distribution of T_i's. The first derivatives of $\log L$ with respect to $\boldsymbol{\theta}$ is

$$\begin{aligned} \frac{\partial \log L}{\partial \theta_k} &= \sum_{i=1}^{n} [\frac{\delta_{1i}\delta_{2i}}{C_{i5} - C_{i6}} \left(\frac{\partial C_{i5}}{\partial \theta_k} - \frac{\partial C_{i6}}{\partial \theta_k} \right) \\ &\quad - \frac{(1-\delta_{1i})\delta_{2i}}{X_{Ri} - X_{Li} - C_{i7}} \frac{\partial C_{i7}}{\partial \theta_k} + \frac{\delta_{1i}(1-\delta_{2i})}{C_{i1} - C_{i2}} \left(\frac{\partial C_{i1}}{\partial \theta_k} - \frac{\partial C_{i2}}{\partial \theta_k} \right)], \\ k &= 1, \ldots, p. \end{aligned} \tag{3.49}$$

The second derivatives of $\log L$ are

$$\begin{aligned}
\frac{\partial^2 \log L}{\partial\theta_k\partial\theta_j} &= \sum_{i=1}^{n} \frac{\delta_{1i}\delta_{2i}}{C_{i5}-C_{i6}}[(\frac{\partial^2 C_{i5}}{\partial\theta_k\partial\theta_j} - \frac{\partial^2 C_{i6}}{\partial\theta_k\partial\theta_j}) \\
&\quad -\frac{1}{C_{i5}-C_{i6}}\left(\frac{\partial C_{i5}}{\partial\theta_k} - \frac{\partial C_{i6}}{\partial\theta_k}\right)\left(\frac{\partial C_{i5}}{\partial\theta_j} - \frac{\partial C_{i6}}{\partial\theta_j}\right)] \\
&\quad -\sum_{i=1}^{n} \frac{(1-\delta_{1i})\delta_{2i}}{X_{Ri}-X_{Li}-C_{i7}}\left[\frac{\partial^2 C_{i7}}{\partial\theta_k\partial\theta_j} + \frac{1}{X_{Ri}-X_{Li}-C_{i7}}\frac{\partial C_{i7}}{\partial\theta_k}\frac{\partial C_{i7}}{\partial\theta_j}\right] \\
&\quad +\sum_{i=1}^{n} \frac{\delta_{1i}(1-\delta_{2i})}{C_{i1}-C_{i2}}[\frac{\partial^2 C_{i1}}{\partial\theta_k\partial\theta_j} - \frac{\partial^2 C_{i2}}{\partial\theta_k\partial\theta_j} \\
&\quad -\frac{1}{C_{i1}-C_{i2}}\left(\frac{\partial C_{i1}}{\partial\theta_k} - \frac{\partial C_{i2}}{\partial\theta_k}\right)\left(\frac{\partial C_{i1}}{\partial\theta_j} - \frac{\partial C_{i2}}{\partial\theta_j}\right)], \qquad (3.50) \\
k, j &= 1, \ldots, p.
\end{aligned}$$

In particular, for the piecewise constant hazard model, the first derivatives of $\log L$ with respect to the ρ_k's are

$$\begin{aligned}
\frac{\partial \log L}{\partial\rho_k} &= \sum_{i=1}^{n}[\frac{\delta_{1i}\delta_{2i}}{C_{i5}-C_{i6}}\left(\frac{\partial C_{i5}}{\partial\rho_k} - \frac{\partial C_{i6}}{\partial\rho_k}\right) \\
&\quad -\frac{(1-\delta_{1i})\delta_{2i}}{X_{Ri}-X_{Li}-C_{i7}}\frac{\partial C_{i7}}{\partial\rho_k} + \frac{\delta_{1i}(1-\delta_{2i})}{C_{i1}-C_{i2}}\left(\frac{\partial C_{i1}}{\partial\rho_k} - \frac{\partial C_{i2}}{\partial\rho_k}\right)], \qquad (3.51) \\
k &= 1, \ldots, r.
\end{aligned}$$

The second derivatives of $\log L$ are

$$\begin{aligned}
\frac{\partial^2 \log L}{\partial\rho_k\partial\rho_j} &= \sum_{i=1}^{n} \frac{\delta_{1i}\delta_{2i}}{C_{i5}-C_{i6}}[(\frac{\partial^2 C_{i5}}{\partial\rho_k\partial\rho_j} - \frac{\partial^2 C_{i6}}{\partial\rho_k\partial\rho_j}) \\
&\quad -\frac{1}{C_{i5}-C_{i6}}\left(\frac{\partial C_{i5}}{\partial\rho_k} - \frac{\partial C_{i6}}{\partial\rho_k}\right)\left(\frac{\partial C_{i5}}{\partial\rho_j} - \frac{\partial C_{i6}}{\partial\rho_j}\right)] \\
&\quad -\sum_{i=1}^{n} \frac{(1-\delta_{1i})\delta_{2i}}{X_{Ri}-X_{Li}-C_{i7}}\left[\frac{\partial^2 C_{i7}}{\partial\rho_k\partial\rho_j} + \frac{1}{X_{Ri}-X_{Li}-C_{i7}}\frac{\partial C_{i7}}{\partial\rho_k}\frac{\partial C_{i7}}{\partial\rho_j}\right]
\end{aligned}$$

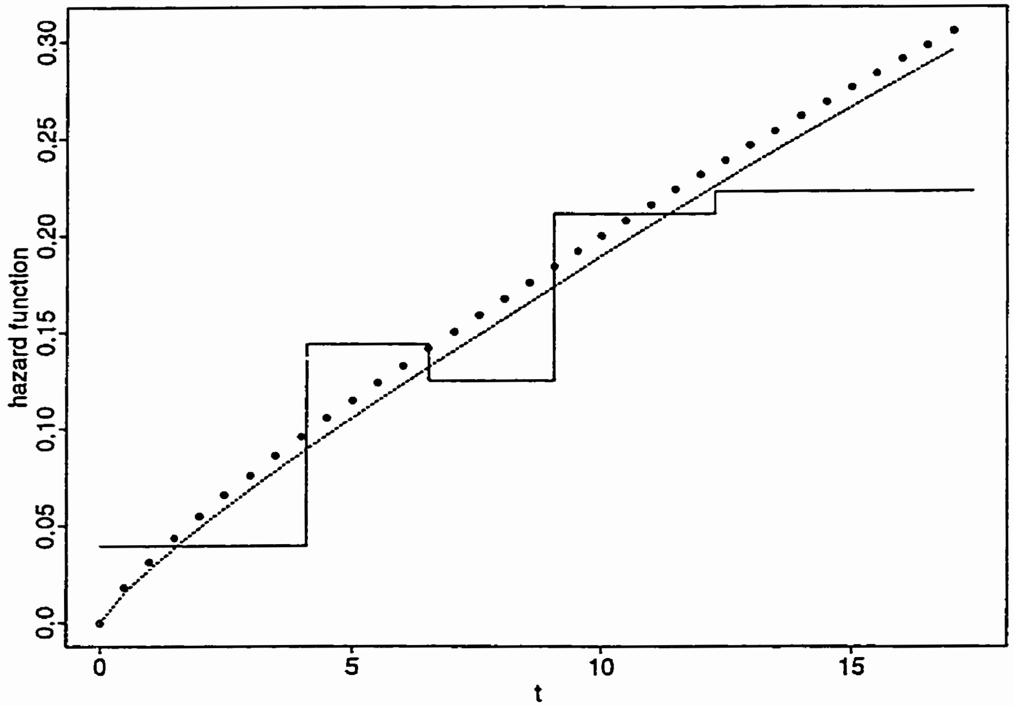

Figure 3.2: Estimates of the true Weibull hazard function (dotted curve) by a piecewise constant model (solid line) and by a Weibull model (dashed curve), for a simulated example in Case 1.

$$
\begin{aligned}
&+\sum_{i=1}^{n}\frac{\delta_{1i}(1-\delta_{2i})}{C_{i1}-C_{i2}}\Big[\frac{\partial^2 C_{i1}}{\partial\rho_k\partial\rho_j}-\frac{\partial^2 C_{i2}}{\partial\rho_k\partial\rho_j}\\
&-\frac{1}{C_{i1}-C_{i2}}\left(\frac{\partial C_{i1}}{\partial\rho_k}-\frac{\partial C_{i2}}{\partial\rho_k}\right)\left(\frac{\partial C_{i1}}{\partial\rho_j}-\frac{\partial C_{i2}}{\partial\rho_j}\right)\Big], \qquad (3.52)\\
k,j &= 1,\ldots,r.
\end{aligned}
$$

For each data set, we estimate the hazard function parameters by maximum likelihood, based on a piecewise constant hazard model and on a Weibull hazard model. The number of pieces is five in the piecewise constant hazard model. The

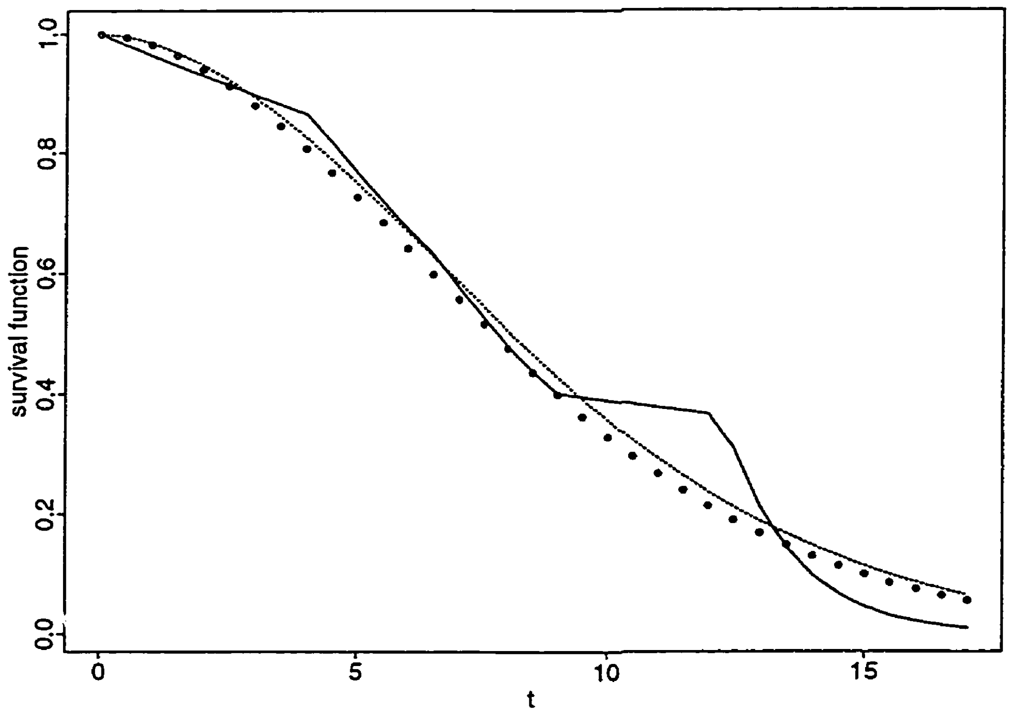

Figure 3.3: Estimates of the true Weibull survival function (dotted curve) by a piecewise constant model (solid line) and by a Weibull model (dashed curve), for the same data used in the previous figure.

values of the a_k's in each case are chosen to be the 20th, 40th, 60th, 80th percentiles of the true distribution of T. The optimization method we used is the downhill simplex method due to Nelder and Mead (Press *et al.*, 1990) and it was programmed in FORTRAN.

The estimates for the survival function of T at the true 5th, 25th, 50th, 75th and 95th percentiles of T are obtained. For piecewise constant hazard models, the survival function value of T at a given point t is estimated by $\widehat{S(t)} = \exp\{-\sum_{k=1}^{r} \hat{\rho}_k u_k(t)\}$, and the asymptotic variance of $\widehat{S(t)}$ is estimated by δ-method, which gives us $var(\widehat{S(t)} = (\widehat{S(t)})^2 u^T(t) cov(\hat{\rho}) u(t)$, where $cov(\hat{\rho})$ is the estimated covariance matrix for $\hat{\rho}$. For Weibull models, let $S_w(t; \alpha, \gamma) = \exp\{-(t/\gamma)^\alpha\}$ be the survival function of a Weibull distribution with parameter α and γ, and let

$$c_w(t; \alpha, \gamma) = (\partial S_w(t)/\partial\alpha, \partial S_w(t)/\partial\gamma)' = ((t/\gamma)^\alpha \log(t/\gamma), -(\alpha/\gamma)(t/\gamma)^\alpha)'.$$

The survival function value of T at a given point t is estimated by $\hat{S}(t) = S_w(t; \hat{\alpha}, \hat{\gamma})$, and the asymptotic variance of $\hat{S}(t)$ is estimated by $var(\hat{S}(t)) = (\hat{S}_w(t))^2 \hat{c}' cov(\hat{\alpha}, \hat{\gamma}) \hat{c}$, where $cov(\hat{\alpha}, \hat{\gamma})$ is the estimated covariance matrix for $(\hat{\alpha}, \hat{\gamma})$. The asymptotic standard error of $\hat{S}(t)$ is obtained by taking square root of $var(\hat{S}(t))$.

Figure 3.2 displays the true Weibull hazard function and its estimates by a piecewise constant model and a Weibull model for a simulated example in Case 1. Figure 3.3 gives the corresponding survival function and its estimates for the same example.

The bias, asymptotic standard error (ASE) and sample standard error (SE) for the estimates of survival functions at the five points for the four settings are given in

Table 3.2: Bias, asymptotic standard error (ASE) and sample standard error (SE) of estimates of survival function for T at five points based on 500 simulations, in Case 1: Weibull distribution, $\omega = 1$, $B = 17.99$

P-C model

t	S(t)	Bias ($\times 10^3$)	ASE ($\times 10^3$)	SE ($\times 10^3$)
1.8116	0.95	-36.5	19.6	18.6
4.7216	0.75	4.86	38.2	35.8
7.6960	0.5	-0.417	45.9	43.2
11.311	0.25	0.125	40.7	39.8
17.3548	0.05	6.38	25.9	27.0

Weibull model

t	S(t)	Bias ($\times 10^3$)	ASE ($\times 10^3$)	SE ($\times 10^3$)
1.8116	0.95	-0.193	14.6	13.8
4.7216	0.75	1.27	36.0	34.0
7.6960	0.5	0.951	40.9	39.0
11.311	0.25	-0.439	36.7	36.1
17.3548	0.05	0.727	19.4	19.7

Table 3.2 to Table 3.5. The bias is the sample mean of the 500 individual differences between $\hat{S(t)}$ and $S(t)$. The ASE is the sample mean of 500 individual standard errors based on the asymptotic theory for the maximum likelihood estimator, using the inverse of the observed information matrix (the second derivative matrix of log-likelihood with respect to parameters, evaluated at the estimates). The SE is the sample standard error of the 500 estimates of $S(t)$.

By comparing the results for piecewise constant hazard models and Weibull models in Table 3.2 and Table 3.3, we can see that when the data come from a Weibull model, the estimates given by Weibull models have smaller bias and

Table 3.3: Bias, asymptotic standard error (ASE) and sample standard error (SE) of estimates of survival function for T at five points based on 500 simulations, in Case 2: Weibull distribution, $\omega = 6$, $B = 17.99$

P-C model

t	S(t)	Bias ($\times 10^3$)	ASE ($\times 10^3$)	SE ($\times 10^3$)
1.8116	0.95	-24.1	33.7	34.1
4.7216	0.75	18.7	46.5	46.0
7.6960	0.5	-9.03	58.5	54.4
11.311	0.25	3.65	59.9	55.8
17.3548	0.05	5.37	38.5	35.4

Weibull model

t	S(t)	Bias ($\times 10^3$)	ASE ($\times 10^3$)	SE ($\times 10^3$)
1.8116	0.95	-0.606	18.2	17.8
4.7216	0.75	1.87	42.7	41.3
7.6960	0.5	1.79	45.1	43.4
11.311	0.25	-0.509	38.7	37.6
17.3548	0.05	0.698	21.7	21.4

standard errors, because the Weibull model is the true model for the data. In Table 3.2, the relative difference of bias (and standard errors) between the Weibull models and the piecewise constant models is not very big for estimation of $S(t)$ in the middle range (from 25th percentile to 75th percentile of the true model), but it is bigger for estimation of $S(t)$ at the two ends (for 5th percentile and 95th percentile). In Table 3.3, where interval censoring of I is more severe, the relative difference of bias (and standard errors) between the Weibull models and the piecewise constant models is bigger for estimation of $S(t)$ in the whole range. These results confirm that using a fully parametric model is more favorable than a piecewise constant

Table 3.4: Bias, asymptotic standard error (ASE) and sample standard error (SE) of estimates of survival function for T at five points based on 500 simulations, in Case 3: log-logistic distribution, $\omega = 1, B = 20.507$

P-C model

t	S(t)	Bias ($\times 10^3$)	ASE ($\times 10^3$)	SE ($\times 10^3$)
2.2665	0.95	-49.9	23.0	21.7
4.7216	0.75	10.7	37.6	35.1
7.3080	0.5	2.00	45.8	43.2
11.311	0.25	2.80	40.5	40.0
23.5634	0.05	-5.36	22.7	24.3

Weibull model

t	S(t)	Bias ($\times 10^3$)	ASE ($\times 10^3$)	SE ($\times 10^3$)
2.2665	0.95	-31.9	19.8	15.1
4.7216	0.75	-3.29	35.8	28.1
7.3080	0.5	44.0	40.8	36.2
11.311	0.25	31.2	37.5	40.9
23.5634	0.05	-35.0	8.71	10.5

model when the true parametric model is known, and the purpose is to estimate the survival function values. The piecewise constant models with just 5 pieces here give reasonably good estimates of survival function values in the middle range, but they may not give satisfactory estimates of survival function values for the tails. In particular, piecewise constant models with 5 cut points at the 0th, 20th, 40th, 60th, and 80th percentiles here under-estimate the survival function value at the 5th percentile. A reason is that the estimated hazard $\hat{\rho}_1$ in the first interval from 0 to 20th percentile is close to the average of hazard function in this interval, so $\hat{\rho}_1$ over estimates the hazard in the sub-interval from 0 to 5th percentile (see Figure

Table 3.5: Bias, asymptotic standard error (ASE) and sample standard error (SE) of estimates of survival function for T at five points based on 500 simulations, in Case 4: log-logistic distribution, $\omega = 6, B = 20.507$

P-C model

t	S(t)	Bias ($\times 10^3$)	ASE ($\times 10^3$)	SE ($\times 10^3$)
2.2665	0.95	-28.7	40.4	44.2
4.7216	0.75	30.5	46.6	46.3
7.3080	0.5	-8.24	56.8	53.8
11.311	0.25	0.0364	44.6	44.5
23.5634	0.05	-3.74	26.9	27.4

Weibull model

t	S(t)	Bias ($\times 10^3$)	ASE ($\times 10^3$)	SE ($\times 10^3$)
2.2665	0.95	-45.7	26.3	23.3
4.7216	0.75	-22.5	42.6	37.2
7.3080	0.5	29.8	44.8	40.7
11.311	0.25	30.6	38.9	40.1
23.5634	0.05	-30.5	11.6	12.9

3.2). We note that these problems can to some extent be overcome by using more pieces and smoothing for the piecewise model, as described in Chapter 4.

The results in Tables 3.4 and 3.5 are based on data generated from the log-logistic model for T; the Weibull and piecewise constant hazard models are still used to fit the data. They show that when we do not know which distribution the data come from, the piecewise constant model gives better estimates of the survival function values in the middle range than a wrongly specified fully parametric model. Both the piecewise constant model and the fully parametric model do not really give satisfactory estimates of survival function values for the left tail in these cases.

We thus see that the piecewise constant models are more robust against the model misspecification. To get good estimates of survival function values in the left tail, we need more pieces and smoothing in the piecewise constant models, which we discuss in the next chapter.

By comparing results in Table 3.2 against these in Table 3.3, and results in Table 3.4 against these in Table 3.5, we can see that when there are more frequent examinations for the initiating event, the estimates from both Weibull models and piecewise constant models are less variable, as we would expect. The bias of estimates in the middle range of T given by the piecewise constant models increases when the number of the periodic examination times for the first event changes from 6 to 1. It suggests that piecewise constant models with a small number of pieces and no smoothing are more sensitive to the heavy interval censoring than a correctly specified fully parametric model.

3.6 Example: Toronto Sexual Contact Study

In this section, we apply the methods discussed in previous sections to the Toronto Sexual Contact Study, described in Chapter 1. The data are doubly censored. The HIV infection times for most subjects were interval-censored; the AIDS diagnosis times were left truncated by the dates of enrollment since only men who had not been diagnosed with AIDS were eligible for entry into the study; and the AIDS diagnosis times for 110 subjects were right censored by the end of follow-up.

We set up the indicator variables δ_{1i} and δ_{2i}, where $\delta_{1i} = 1$ if the subject was diagnosed with AIDS during the study, 0 otherwise, $\delta_{2i} = 1$ if the subject was HIV positive at the time of enrollment, 0 otherwise. The data can then be divided into four groups. The first group ($n_1 = 15$, with $\delta_{1i} = 0$, $\delta_{2i} = 0$) consists of subjects who became HIV positive during the study and were AIDS-free at the end of follow-up; the second group ($n_2 = 1$, with $\delta_{1i} = 1$, $\delta_{2i} = 0$) consists of subjects who became HIV positive during the study and were diagnosed with AIDS during the study; the third group ($n_3 = 95$, with $\delta_{1i} = 0$, $\delta_{2i} = 1$) consists of subjects who were HIV positive at enrollment and were AIDS-free at the end of follow-up; the fourth group ($n_4 = 48$, with $\delta_{1i} = 1$, $\delta_{2i} = 1$) consists of subjects who were HIV positive at enrollment and were diagnosed with AIDS during the study.

We focus on getting an estimate for the distribution of the incubation period from HIV infection to AIDS diagnosis. We apply the piecewise-constant hazard models to the data. That is, we assume that T_i, the time between the HIV infection and AIDS diagnosis has a piecewise constant hazard, as discussed in the early sections of this chapter; the distribution of the time to HIV infection is assumed to be known and I_i and T_i are independent. Under these assumptions, the likelihood contribution from the ith subject is

$$\begin{aligned} L_i &= S_i(J_{Li} - I_i)^{(1-\delta_{1i})(1-\delta_{2i})} g_i(J_{Li} - I_i)^{\delta_{1i}(1-\delta_{2i})} \\ &\quad \left\{ \frac{\int_{X_{Li}}^{X_{Ri}} w_i^*(x) S_i(J_{Li} - x)dx}{\int_{X_{Li}}^{X_{Ri}} w_i^*(x) S_i(E_i - x)dx} \right\}^{(1-\delta_{1i})\delta_{2i}} \\ &\quad \left\{ \frac{\int_{X_{Li}}^{X_{Ri}} w_i^*(x) g_i(J_{Li} - x)dx}{\int_{X_{Li}}^{X_{Ri}} w_i^*(x) S_i(E_i - x)dx} \right\}^{\delta_{1i}\delta_{2i}}, \end{aligned} \tag{3.53}$$

where S_i and g_i are the survival function and the density function of T_i, and $w_i^*(x) = w_i(x)/\int_{X_{Li}}^{X_{Ri}} w_i(u)du$ is the conditional density of I_i on interval $[X_{Li}, X_{Ri}]$.

If we assume that T_i's follow a common distribution $S(t)$, and w_i^* is uniform over interval $[X_{Li}, X_{Ri}]$, L_i can be written as

$$L_i = S(J_{Li} - I_i)^{(1-\delta_{1i})(1-\delta_{2i})} g(J_{Li} - I_i)^{\delta_{1i}(1-\delta_{2i})}$$
$$\left\{\frac{IS(J_{Li} - X_{Li}, 0) - IS(J_{Li} - X_{Ri}, 0)}{IS(E_i - X_{Li}, 0) - IS(E_i - X_{Ri}, 0)}\right\}^{(1-\delta_{1i})\delta_{2i}}$$
$$\left\{\frac{S(J_{Li} - X_{Ri}) - S(J_{Li} - X_{Li})}{IS(E_i - X_{Li}, 0) - IS(E_i - X_{Ri}, 0)}\right\}^{\delta_{1i}\delta_{2i}}. \quad (3.54)$$

For the special case of $X_{Li} = X_{Ri}$ in Groups Three and Four, we treat I_i as observed at X_{Li}, and replace the corresponding likelihood contribution by $S(J_{Li} - X_{Li})/S(E_i - X_{Li})$ in Group Three and by $g(J_{Li} - X_{Li})/S(E_i - X_{Li})$ in Group Four. The log likelihood can be written as $\log L = \sum_{i=1}^{n} \log L_i$.

A piecewise constant model with 6 pieces was fitted to the data without any covariates. The cutoff points are $0.0, 500.0, 1000.0, 1500.0, 2200.0, 2700.0$. The maximum likelihood estimates of the hazards and estimated standard errors based on the observed information matrix are listed in Table (3.6). A Weibull model with survivor function $S(t; \alpha, \gamma) = exp\{-(t/\gamma)^\alpha\}$ was also fitted to the data. The maximum likelihood estimates are $\hat{\alpha} = 1.7724$, $\hat{\gamma} = 3715.1$. The corresponding estimated standard errors are $se(\hat{\alpha}) = 0.2902$, $se(\hat{\gamma}) = 365.0$.

Table 3.6: Estimates and standard errors of the hazards in a piecewise constant hazard model

Interval	(0, 500]	(500,1000]	(1000,1500]	(1500,2200]	(2200,2700]	(2700, ∞]
$\hat{\rho}_k$ ($\times 10^{-4}$)	0.4597	1.840	1.492	2.841	2.142	7.855
$se(\hat{\rho}_k)$ ($\times 10^{-4}$)	0.4544	0.6793	0.6481	0.7939	1.315	3.335

Figure 3.4 gives the estimate of CDF of T by the Weibull model (solid curve) and the estimate by the above piecewise constant model. We notice the discrepancy in the right tail, but observe that there are few subjects with observed t_i's which could be larger than 2500 days. The two models give similar estimates of the median incubation time.

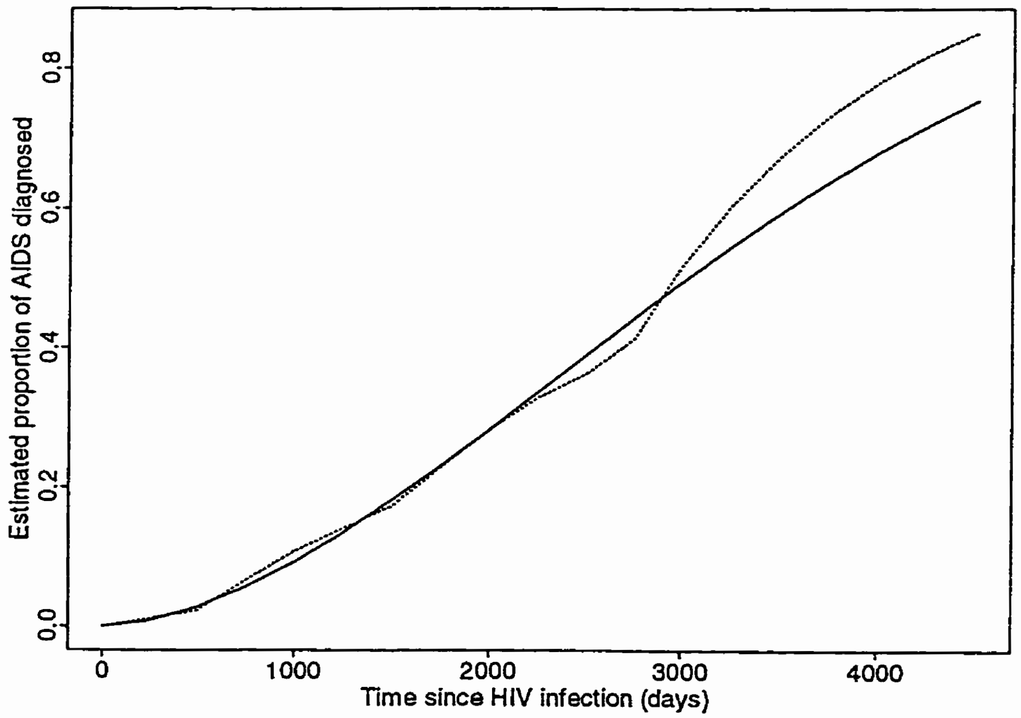

Figure 3.4: The solid line is the estimated CDF of T based on a Weibull model; the dotted line is the estimated CDF of T based on a piecewise constant hazard model with 6 pieces.

3.7 Discussion

In this chapter, we have discussed models with piecewise constant hazard functions for current status data and doubly censored data. Maximum likelihood estimation of the survival function of the induction time and covariate effects is developed. The use of these methods is illustrated by simulated examples and an example from an AIDS study. The performance of these methods are also demonstrated by a simulation study.

In the simulated examples for the standard current status data, we have shown that the piecewise constant models provide reasonable estimates of the regression coefficient and the corresponding estimated standard errors. The estimates of the regression coefficient and the corresponding estimated standard errors are robust to the number of pieces used in the piecewise constant models.

The simulation study in Section 3.5 assumes a Weibull or log-logistic distribution for the induction time and no covariates involved. The models with a piecewise constant hazard function or a Weibull hazard function were fitted. Results show that use of piecewise constant hazard functions with five to ten pieces can provide good estimates of the survival functions when time is not too small, no matter what is the true distribution of the induction time. On the other hand, the Weibull models do not do well in estimating the survival function when the induction times are from a log-logistic distribution. Therefore the piecewise constant models are robust to the distributional form of the induction time and are recommended to use when we have doubt about the distributional form of the induction time. The estimates of the standard errors can be constructed using the observed information matrices of the parameter estimates.

A more thorough investigation can be carried out by assessing the performance of the piecewise constant models in the doubly-censored data when covariates are present. A known distribution other than the uniform distribution can also be considered for the initiating event I.

3.8 Appendix

For the proportional hazard model with piecewise constant baseline hazard in Section 3.2.2, Section 3.3.2 and Section 3.4, the definitions of d_i and $S(c|z_i)$ can be found just before formula 3.16, and $IS(c, z_i)$ is given in formula (3.16). The derivatives of d_i and $S(t|z_i)$ with respect to ρ_k's and β_k's are the following:

$$d_i = \exp(z_i'\beta), \tag{3.55}$$

$$\frac{\partial d_i}{\partial \beta_k} = d_i z_{ik}, \tag{3.56}$$

$$\frac{\partial^2 d_i}{\partial \beta_k \beta_j} = d_i z_{ik} z_{ij}, \tag{3.57}$$

$$\frac{\partial S(t|z_i)}{\partial \rho_k} = -u_k(t) d_i S(t|z_i), \tag{3.58}$$

$$\frac{\partial S(t|z_i)}{\partial \beta_j} = -d_i z_{ij} H_0(t) S(t|z_i), \tag{3.59}$$

$$\frac{\partial^2 S(t|z_i)}{\partial \rho_k \partial \rho_j} = u_k(t) u_j(t) d_i^2 S(t|z_i), \tag{3.60}$$

$$\frac{\partial^2 S(t|z_i)}{\partial \beta_k \partial \beta_j} = -d_i z_{ik} z_{ij} H_0(t) S(t|z_i) + d_i^2 z_{ik} z_{ij} H_0^2(t) S(t|z_i), \tag{3.61}$$

$$\frac{\partial^2 S(t|z_i)}{\partial \rho_k \partial \beta_j} = -d_i z_{ij} u_k(t) S(t|z_i) + d_i^2 z_{ij} u_k(t) H_0(t) S(t|z_i). \tag{3.62}$$

The derivatives of $IS(c, z_i)$ with respect to ρ_k's and β_k's are the following:

$$\begin{aligned}
\frac{\partial IS(c,z_i)}{\partial \rho_k} &= -\sum_{q=1}^{r} \rho_q^{-1} u_k(a_q) S(a_q|z_i)[1 - exp(-\rho_q d_i u_q(c))] \\
&\quad -\rho_k^{-2} d_i^{-1} S(a_k|z_i)[1 - exp(-\rho_k d_i u_k(c))] \\
&\quad +\rho_k^{-1} u_k(c) S(a_k|z_i) exp(-\rho_k d_i u_k(c)), \qquad (3.63) \\
\frac{\partial IS(c,z_i)}{\partial \beta_j} &= -z_{ij} IS(c, z_i) + z_{ij} \sum_{q=1}^{r} u_q(c) S(a_q|z_i) exp(-\rho_q d_i u_q(c)) \\
&\quad - \sum_{q=1}^{r} \rho_q^{-1} z_{ij} H_0(a_q) S(a_q|z_i)[1 - exp(-\rho_q u_q(c) d_i)], \qquad (3.64) \\
\frac{\partial^2 IS(c,z_i)}{\partial \rho_k^2} &= \sum_{q=1}^{r} \rho_q^{-1} u_k^2(a_q) d_i S(a_q|z_i)[1 - exp(-\rho_q d_i u_q(c))] \\
&\quad +2\rho_k^{-3} d_i^{-1} S(a_k|z_i)[1 - exp(-\rho_k d_i u_k(c))] \\
&\quad -2\rho_k^{-2} u_k(c) S(a_k|z_i) exp(-\rho_k d_i u_k(c)) \\
&\quad -\rho_k^{-1} d_i u_k^2(c) S(a_k|z_i) exp(-\rho_k d_i u_k(c)), \qquad (3.65) \\
\frac{\partial^2 IS(c,z_i)}{\partial \rho_k \partial \rho_j} &= \sum_{q=1}^{r} \rho_q^{-1} u_k(a_q) u_j(a_q) d_i S(a_q|z_i)[1 - exp(-\rho_q d_i u_q(c))] \\
&\quad +\rho_j^{-2} u_k(a_j) S(a_j|z_i)[1 - exp(-\rho_j d_i u_j(c))] \\
&\quad +\rho_k^{-2} u_j(a_k) S(a_k|z_i)[1 - exp(-\rho_k d_i u_k(c))] \\
&\quad -\rho_j^{-1} u_k(a_j) u_j(c) d_i S(a_j|z_i) exp(-\rho_j d_i u_j(c)) \\
&\quad -\rho_k^{-1} u_j(a_k) u_k(c) d_i S(a_k|z_i) exp(-\rho_k d_i u_k(c)), \qquad (3.66)
\end{aligned}$$

for $k \neq j$,

$$\begin{aligned}
\frac{\partial^2 IS(c,z_i)}{\partial \beta_k \beta_j} &= +z_{ij} z_{ik} IS(c, z_i) \\
&\quad +z_{ij} z_{ik} \sum_{q=1}^{r} \rho_q^{-1} H_0(a_q) S(a_q|z_i)[1 - exp(-\rho_q d_i u_q(c))] \\
&\quad -z_{ij} z_{ik} \sum_{q=1}^{r} u_q(c) S(a_q|z_i) exp(-\rho_q d_i u_q(c))
\end{aligned}$$

$$
\begin{aligned}
& +d_i z_{ij} z_{ik} \sum_{q=1}^{r} \rho_q^{-1} H_0^2(a_q) S(a_q|z_i)[1 - exp(-\rho_q d_i u_q(c))] \\
& -2 d_i z_{ij} z_{ik} \sum_{q=1}^{r} u_q(c) H_0(a_q) S(a_q|z_i) exp(-\rho_q d_i u_q(c)) \\
& -d_i z_{ij} z_{ik} \sum_{q=1}^{r} \rho_q u_q^2(c) S(a_q|z_i) exp(-\rho_q d_i u_q(c)), \qquad (3.67)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 IS(c, z_i)}{\partial \rho_k \beta_j} = \; & d_i z_{ij} \sum_{q=1}^{r} \rho_q^{-1} u_k(a_q) H_0(a_q) S(a_q|z_i)[1 - exp(-\rho_q d_i u_q(c))] \\
& -d_i z_{ij} \sum_{q=1}^{r} u_k(a_q) u_q(c) S(a_q|z_i) exp(-\rho_q d_i u_q(c)) \\
& +\rho_k^{-2} d_i^{-1} z_{ij} S(a_k|z_i)[1 - exp(-\rho_k d_i u_k(c))] \\
& +\rho_k^{-1}(\rho^{-1} - 1) z_{ij} H_0(a_k) S(a_k|z_i)[1 - exp(-\rho_k d_i u_k(c))] \\
& -(\rho_k^{-1} + d_i u_k(c)) z_{ij} u_k(c) S(a_k|z_i) exp(-\rho_k d_i u_k(c)). \qquad (3.68)
\end{aligned}
$$

For a Weibull model discussed in Section 3.4.3, $IS(t, z_i)$ is given in formula (3.45), and $S(t|z_i)$ and d_i are defined just before formula (3.45). The derivatives of $S(t|z_i)$ and $IS(t, z_i)$ with respect to α and γ are the following:

$$
\begin{aligned}
\frac{\partial S(t|z_i)}{\partial \alpha} &= -S(t|z_i) H_0(t) d_i \log(t/\gamma), && (3.69) \\
\frac{\partial S(t|z_i)}{\partial \gamma} &= S(t|z_i) d_i \alpha/\gamma H_0(t), && (3.70) \\
\frac{\partial S(t|z_i)}{\partial \beta_k} &= -S(t|z_i) H_0(t) d_i z_{ik}, && (3.71) \\
\frac{\partial^2 S(t|z_i)}{\partial \alpha^2} &= S(t|z_i) H_0(t) d_i (\log(t/\gamma))^2 (d_i H_0(t) - 1), && \\
\frac{\partial^2 S(t|z_i)}{\partial \alpha \partial \gamma} &= S(t|z_i) H_0(t) d_i/\gamma [1 + \alpha log(t/\gamma)(1 - d_i H_0(t))], && (3.72) \\
\frac{\partial^2 S(t|z_i)}{\partial \gamma^2} &= S(t|z_i) d_i \alpha/\gamma^2 H_0(t)[d_i \alpha H_0(t) - \alpha - 1], && (3.73)
\end{aligned}
$$

$$\frac{\partial^2 S(t|z_i)}{\partial\gamma\beta_k} = S(t|z_i)H_0(t)\alpha/\gamma d_i z_{ik}[1 - d_i H_0(t)], \tag{3.74}$$

$$\frac{\partial^2 S(t|z_i)}{\partial\alpha\partial\beta_k} = S(t|z_i)H_0(t)\log(t/\gamma)d_i z_{ik}[d_i H_0(t) - 1], \tag{3.75}$$

$$\frac{\partial^2 S(t|z_i)}{\partial\beta_k\partial\beta_j} = S(t|z_i)H_0(t)d_i z_{ik} z_{ij}[d_i H_0(t) - 1], \tag{3.76}$$

$$\begin{aligned}
\frac{\partial IS(t, z_i)}{\partial\alpha} &= -\int_0^t S(x|z_i)H_0(x)d_i \log(x/\gamma)dx \\
&= -d_i\gamma\int_0^{t/\gamma} exp(-d_i u^\alpha)u^\alpha \log u du, \qquad (3.77)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial IS(t, z_i)}{\partial\gamma} &= \int_0^t S(x|z_i)H_0(x)d_i\alpha/\gamma dx \\
&= d_i^{-1/\alpha}\Gamma(1/\alpha+1)P(1/\alpha+1, (t/\gamma)^\alpha d_i) \\
&= d_i^{-1/\alpha}[\alpha^{-1}\Gamma(1/\alpha)P(1/\alpha, (t/\gamma)^\alpha) - (t/\gamma)S(t|z_i)], \qquad (3.78)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 IS(t, z_i)}{\partial\alpha^2} &= \int_0^t S(x|z_i)H_0(x)d_i(log(x/\gamma))^2[d_i H_0(x) - 1]dx \\
&= d_i\gamma\int_0^{t/\gamma} exp(-d_i u^\alpha)u^\alpha(logu)^2[d_i u^\alpha - 1]du, \qquad (3.79)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 IS(t, z_i)}{\partial\alpha\partial\gamma} &= \int_0^t S(x|z_i)H_0(x)[1 + \alpha log(x/\gamma)(1 - d_i H_0(x))]dx \\
&= d_i\int_0^{t/\gamma} exp(-u^\alpha d_i)u^\alpha[1 + \alpha\log u(1 - d_i u^\alpha)]du, \qquad (3.80)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 IS(t, z_i)}{\partial\gamma^2} &= d_i\alpha/\gamma^2\int_0^t S(x|z_i)H_0(x)[d_i\alpha H_0(x) - 1]dx \\
&= \gamma^{-1}d_i^{-1/\alpha}[\Gamma(1/\alpha+2)P(1/\alpha+2, (t/\gamma)^\alpha d_i)\alpha \\
&\quad -(\alpha+1)\Gamma(1/\alpha+1)P(1/\alpha+1, (t/\gamma)^\alpha d_i)] \\
&= -d_i\alpha/\gamma(t/\gamma)^{1+\alpha}S(t|z_i). \qquad (3.81)
\end{aligned}$$

Chapter 4

Smoothing in Estimation of Rate and Hazard Functions

4.1 Smoothing using a roughness-penalized likelihood approach

In previous chapters we have discussed methods using piecewise constant intensity, rate or hazard functions for interval-grouped recurrent event data and doubly-censored data. We have noticed that for the purpose of estimating covariate effects and mean function or survival function, a small number of pieces can usually do well. However, if our goal is to estimate intensity, rate or hazard functions, a larger number of pieces is normally required. In addition, estimates of survival, mean or cumulative hazard functions for small and large t values is often not very good when only a few pieces are used. Using a large number of parameters makes the estimation problem ill-posed (O'Sullivan, 1986) in the sense that small changes in the data may lead to large changes in the estimate. It also yields very "wiggly" estimates, and the likelihood may often not have a unique maximum. A common strategy to overcome the difficulties is to maximize a roughness-penalized likelihood (see, e.g., Bacchetti, 1990; Bacchetti and Jewell, 1991; Green and Silverman, 1994). The general form of a penalized log-likelihood is

$$\log(L) - (1/2)\zeta R, \quad \zeta \geq 0, \tag{4.1}$$

where L is the likelihood based on observed data, R is a penalty function representing the roughness of the hazard function, and ζ is a tuning constant that determines the relative importance of L and R. For models with piecewise constant intensity, rate or hazard functions, R is usually chosen to be the the sum of squared second differences among the piecewise constant intensity, rate or hazard:

$$R(\rho) = \sum_{k=1}^{r-2} (\rho_k - 2\rho_{k+1} + \rho_{k+2})^2 = \boldsymbol{\rho}^T K \boldsymbol{\rho}, \tag{4.2}$$

where ρ_k's are piecewise constant intensity, rate or hazard values, and K is the corresponding penalty matrix. The penalized likelihood technique balances its fit to the data against the prior knowledge that rougher estimates are less plausible (Bacchetti, 1990), and it can be interpreted from a Bayesian viewpoint.

The estimate produced by maximizing a roughness-penalized likelihood is called a maximum penalized likelihood estimate (MPLE). There are different ways to compute MPLE. For example, Bacchetti (1990) proposed using an EM algorithm to maximize a roughness-penalized likelihood in estimating HIV infection rates and AIDS incubation distribution in a discrete time setting. He also used bootstrap simulations to get confidence intervals for infection rates. As pointed out by Segal *et al.* (1994), variance estimates for maximum penalized likelihood estimates (MPLEs) can be obtained by treating the penalized likelihood as a usual likelihood and inverting the observed information matrices. These estimates are derived by Silverman (1985) using a Bayesian model. He showed that these estimates are the posterior variance matrices for multivariate normal data. Segal *et al.* (1994) also developed a procedure for obtaining these variance matrices when the MPLEs are obtained through an EM algorithm.

The tuning constant ζ can be chosen by visually examining the smoothness of the estimates over a plausible range of ζ values (e.g., Bacchetti, 1990; Fusaro *et al.*, 1996), or by an automatic procedure based on some generalized cross-validation criterion (e.g., Marschner, 1997; Joly *et al.*, 1998).

The above penalized likelihood technique is not the only technique for smoothing. Other techniques include kernel smoothing (e.g., Staniswalis *et al.*, 1997), weighted locally linear smoothers (e.g., Cleveland, 1979), or spline functions (e.g., Joly *et al.*, 1998). However, we will focus on the penalized likelihood technique in this chapter. In the rest of this chapter, we discuss the application of the penalized likelihood technique to produce smoothed estimates of intensity, rate or hazard functions for recurrent event data, current status data and doubly-censored data.

4.2 Smoothing in estimation of intensity or rate function for interval-grouped recurrent event data

For the interval-grouped recurrent event data, we assume that the event process is a mixed Poisson process as in Chapter 2. The intensity for event recurrence is ρ_k for time interval $A_k = (a_{k-1}, a_k]$, where A_k's are of equal length. We choose R to be the sum of squared second differences among rates:

$$R(\boldsymbol{\rho}) = \sum_{k=1}^{r-2} (\rho_k - 2\rho_{k+1} + \rho_{k+2})^2 = \boldsymbol{\rho}^T K \boldsymbol{\rho}, \tag{4.3}$$

where K is equal to $W^T W$ with tridiagonal such that $w_{k,k} = 1$, $w_{k,k+1} = -2$, $w_{k,k+2} = 1$.

Therefore the penalized log-likelihood is

$$\begin{aligned} l_{pe}(\boldsymbol{\rho}, \boldsymbol{\beta}, v) &= l(\boldsymbol{\rho}, \boldsymbol{\beta}, v) - (1/2)\zeta R(\boldsymbol{\rho}) \\ &= \sum_{i=1}^{m} \sum_{j} n_{ij} \log \mu_{ij} + n_{i.} \log v + \log \Gamma(n_{i.} + 1/v) - \log \Gamma(1/v) \\ &\quad - (n_{i.} + 1/v) \log(1 + v\mu_{i.}) - (1/2)\zeta \boldsymbol{\rho}^T K \boldsymbol{\rho}. \end{aligned} \tag{4.4}$$

Our goal is to obtain the maximum penalized likelihood estimate (MPLE) of $(\boldsymbol{\rho}, \boldsymbol{\beta}, v)$ by maximizing l_{pe}.

The first derivatives of l_{pe} with respect to parameters are

$$\begin{aligned}
\frac{\partial l_{pe}}{\partial \rho} &= \frac{\partial l}{\partial \rho} - \zeta K \rho, \\
\frac{\partial l_{pe}}{\partial \beta} &= \frac{\partial l}{\partial \beta}, \\
\frac{\partial l_{pe}}{\partial v} &= \frac{\partial l}{\partial v}.
\end{aligned} \tag{4.5}$$

The minus second order derivatives of l_{pe} with respect to parameters are

$$\frac{-\partial^2 l_{pe}}{\partial \rho_k \rho_l} = \frac{-\partial^2 l}{\partial \rho_k \rho_l} + \zeta K_{kl}, \quad k, l = 1, \ldots, r, \tag{4.6}$$

and $\frac{-\partial^2 l_{pe}}{\partial \rho_k \partial \beta_l}$, $\frac{-\partial^2 l_{pe}}{\partial \beta_k \partial \beta_l}$, $\frac{-\partial^2 l_{pe}}{\partial \rho_k \partial v}$, $\frac{-\partial^2 l_{pe}}{\partial \beta_k \partial v}$, $\frac{-\partial^2 l_{pe}}{\partial v^2}$, which are the same as the corresponding minus second order derivatives of the observed log-likelihood.

We apply a two-step algorithm to get MPLE. Suppose the current parameter values are $(\rho^{(0)}, \beta^{(0)}, v^{(0)})$. At Step one we update (ρ, β) by Fisher's scoring method:

$$\begin{aligned}
\left(\rho^{(1)T}, \beta^{(1)T}\right)^T &= \left(\rho^{(0)T}, \beta^{(0)T}\right)^T + \\
& E\{\frac{-\partial^2 l_{pe}}{\partial(\rho^{(0)T}, \beta^{(0)T})\partial(\rho^{(0)T}, \beta^{(0)T})^T}\}^{-1} \frac{\partial l_{pe}}{\partial(\rho^{(0)}, \beta^{(0)})},
\end{aligned} \tag{4.7}$$

and at Step two v is updated by

$$v^{(1)} = v^{(0)} + \{\frac{-\partial^2 l_{pe}}{\partial v^{(0)2}}\}^{-1} \frac{\partial l_{pe}}{\partial v^{(0)}}. \tag{4.8}$$

We iterate between the two steps. The stopping rule is defined as follows: Let

$$\delta_1 = \max_k \frac{|\rho_k^{(1)} - \rho_k^{(0)}|}{\rho_k^{(0)} + 0.001},$$
$$\delta_2 = \max_k \frac{|\beta_k^{(1)} - \beta_k^{(0)}|}{|\beta_k^{(0)}| + 0.001},$$
$$\delta_3 = \frac{|v^{(1)} - v^{(0)}|}{v^{(0)} + 0.001}, \tag{4.9}$$

and δ_4 is the maximum absolute value of the penalized score functions evaluated at the updated parameters. If $\max(\delta_1, \delta_2, \delta_3) \leq \epsilon_1$ and $\delta_4 \leq \epsilon_2$, where ϵ_1, ϵ_2 are small positive numbers, then stop iterations and claim that MPLE is found. Here we use $\epsilon_1 = 1.0 \times 10^{-6}$ and $\epsilon_2 = 1.0 \times 10^{-4}$.

4.2.1 An example

We apply the penalized likelihood approach to the bladder cancer data discussed in Chapter 2. We divided the whole study period $(0, 53]$ into 53 pieces of equal length. Recall that covariate z_i is the treatment indicator; $z_{i1} = 1$ if the patient is in the thiotepa group; $z_{i1} = 0$ if the patient is in the placebo group. z_{i2} is the number of initial tumors present at randomization; z_{i3} is the diameter (in centimeters) of the largest initial tumor. z_2 and z_3 were centered before entering the estimation procedure. Figure 4.1 shows the estimates of the baseline rate function using the penalized likelihood method with different degrees of smoothness. $\zeta = 1.0 \times 10^4$ produces rates with oscillations. $\zeta = 1.0 \times 10^5$ gives rates that are reasonably smooth. $\zeta = 1.0 \times 10^6$ gives rates that have little curvature.

Setting $\zeta = 1.0 \times 10^5$ gives $\hat{\beta} = (-0.9207, 0.3567, 0.0043)$, with standard errors

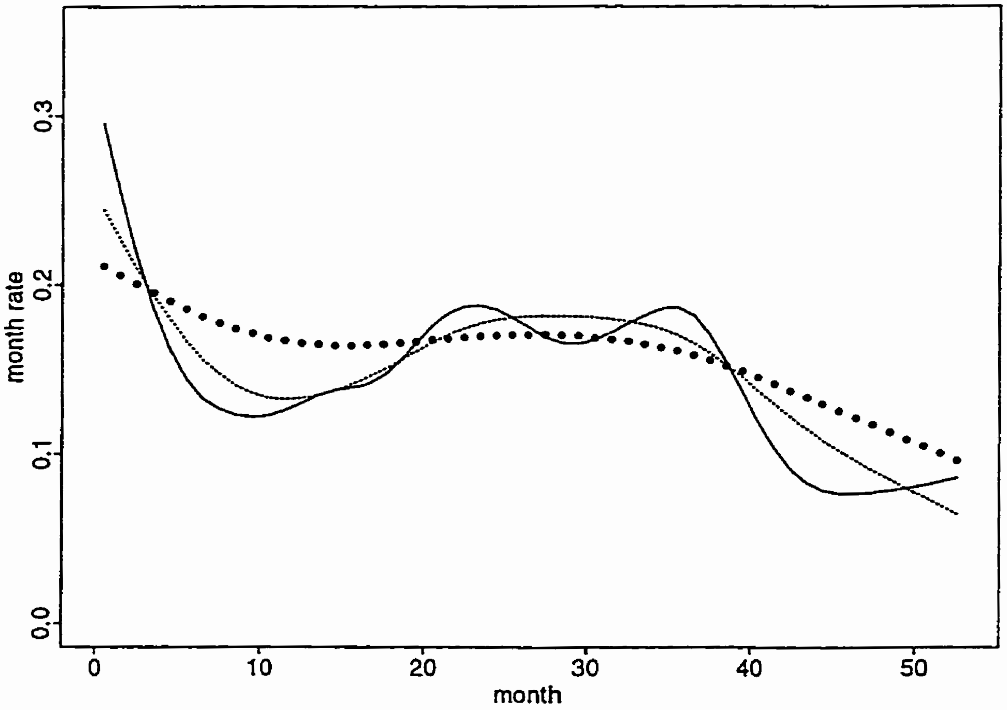

Figure 4.1: Monthly baseline recurrence rates estimated by penalized likelihood method with $\zeta = 1.0 \times 10^4$ (solid curve), $\zeta = 1.0 \times 10^5$ (short dashed curve), and $\zeta = 1.0 \times 10^6$ (bold dots).

$(0.37, 0.105, 0.13)$. The variance of the random effect is estimated as 2.43, with standard error 0.500.

4.3 Smoothing in estimation of hazard functions for current status data and doubly-censored data

We will consider the doubly-censored data in this section, although the procedure for current status data is very similar. The penalized log-likelihood function for doubly-censored data is:

$$l_{pe}(\rho,\beta) = l(\rho,\beta) - (1/2)\zeta\rho^T K\rho, \tag{4.10}$$

where K is the penalty matrix as defined before, $l(\rho,\beta)$ is the log-likelihood based on observed data, ρ is the piecewise hazard. The maximum penalized likelihood estimate (MPLE) of (ρ,β) can be obtained by maximizing l_{pe}. A derivative- free method (e.g., the downhill simplex method of Nelder and Mead, see Press *et al.* (1990)) can be applied to achieve this goal.

4.3.1 A Simulation Study

To assess the performance of the MPLEs in piecewise constant models, we carry out a simulation study for doubly-censored data. The design of the study is the same as the one in Section 3.5, with the same four cases. We fit models with piecewise constant hazards for T. The number of pieces is 25 and the values of endpoints a_k's are chosen to be the percentiles of the true distribution of T that correspond to CDF values $0, 0.036, 0.072, \ldots, 0.864$. The penalized likelihood technique was used to get MPLEs of the hazard function. The tuning constant is set to be 10^4 after trying several different ζ values for a simulated data set in Case 4. For this data set, Figure 4.2 shows the the true log-logistic hazard function and its estimates by piecewise constant models with tuning constant being 10^4 or 10^5, and its estimate by a Weibull model. We can see that the two estimates by piecewise constant models are very close in the middle but slightly different at the tails. Both give a very good approximation to the true hazard function in the range of roughly $(3, 12]$, but are not as good for the tails (where there is little data). The estimated hazard from the Weibull model can not of course match the pattern of the true hazard function.

Figure 4.3 gives the estimates of survival function for the same data set. Clearly, piecewise constant models with both tuning parameters give good estimates to the true survival function, but the Weibull model does not do so well, especially in the middle range of T.

Tables 4.1 to 4.4 give the bias, asymptotic standard error (ASE) and sample standard error (SE) for the estimates of survival functions at five points (5th, 25th,50th, 75th and 95th percentiles of the true distribution) for the four cases. The asymptotic standard errors are the square roots of the the diagonal elements of the estimated asymptotic variance matrix, which is computed as the inverse matrix of the minus second partials of the penalized log likelihood function. The piecewise constant models with smoothing give good estimates for survival functions. We notice that the asymptotic standard errors of the $\hat{S}(t)$'s are larger than the sample standard errors. This suggests that asymptotic standard errors based on the minus second derivative matrices of the penalized log-likelihood are not valid. See the discussion in the following section. We suggest using the bootstrap estimates of the standard errors instead.

Figure 4.4 shows the 95% pointwise confidence interval for the true hazard function for Case 4, based on 0.025th and 0.975th quantiles of individual estimates of ρ_k's in the 500 simulations. The interval covers the true hazard function, except very close to $t = 0$. The interval is quite narrow in the middle, but becomes wider for t greater than 14, indicating a lack of information for the hazard when t is larger than 14.

Table 4.1: Bias, asymptotic standard error (ASE) and sample standard error (SE) of estimates of survival function for T at five points based on 500 simulations, in Case 1: Weibull distribution, $\omega = 1$, $B = 17.99$

P-C model

t	S(t)	Bias ($\times 10^3$)	ASE ($\times 10^3$)	SE ($\times 10^3$)
1.8116	0.95	-3.02	23.1	18.2
4.7216	0.75	1.24	39.8	24.0
7.6960	0.5	0.460	47.3	22.4
11.311	0.25	-1.73	39.9	6.54
17.3548	0.05	3.35	24.5	6.85

4.4 Discussion

In Section 4.3.1, we noticed the discrepancy between the the asymptotic standard errors (ASE) and the sample standard errors (SSE) for $\hat{S}(t)$. Based on Tables 4.1 to 4.4, the ASE is always larger than the SSE and the ratio of ASE over SSE can

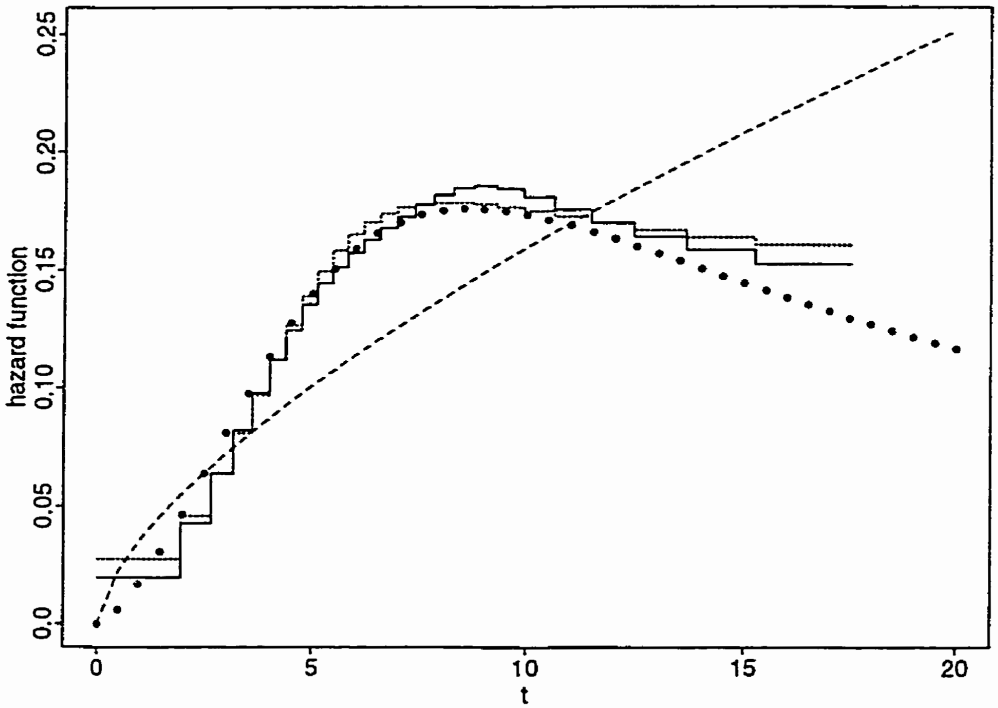

Figure 4.2: Estimates of the true log-logistic hazard function (points) by piecewise constant models with tuning constant $\zeta = 10^4$ (solid line) or $\zeta = 10^5$ (dotted line), and by a Weibull model (dashed curve), for a simulated example in Case 4.

be as high as 5. It indicates that the ASE based on the minus second derivative matrix of the penalized log-likelihood function is not valid.

An alternative approach is to compute a bootstrap estimate of standard errors (e.g., Efron and Tibshirani, 1993). Suppose we have a sample of size n, $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$ and we have computed an estimate of $S(t)$ for some time t. The estimate is denoted as $\hat{S}(t)$. Now we want to compute the bootstrap standard error of $\hat{S}(t)$. The steps are the following. First we generate M independent bootstrap

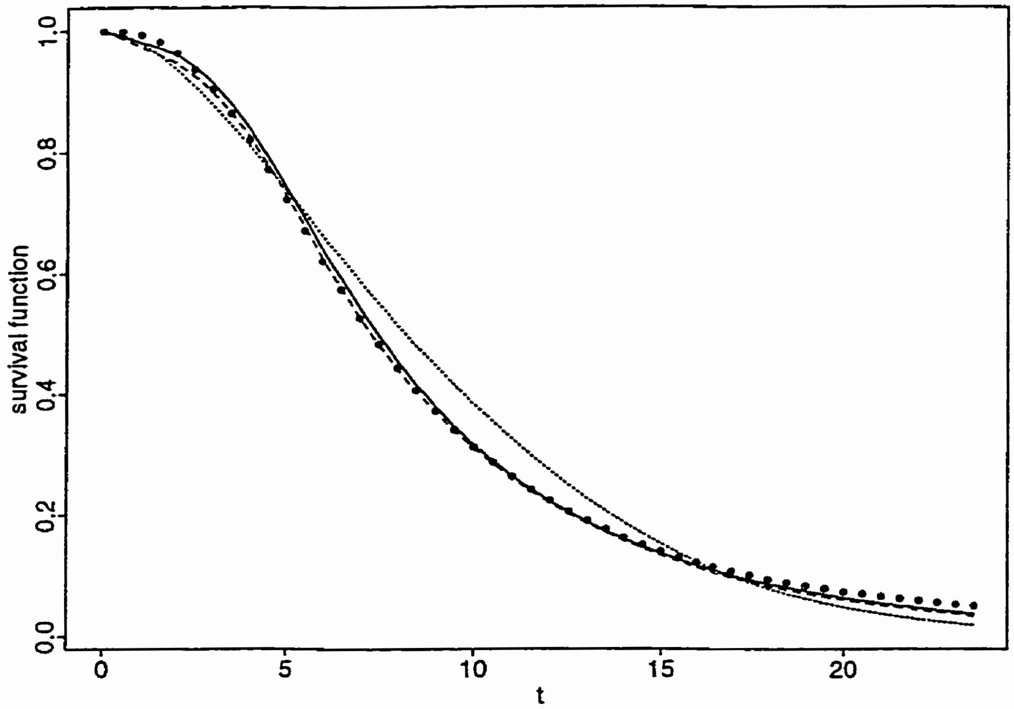

Figure 4.3: Estimates of the true log-logistic survival function (points) by piecewise constant models with tuning constant $\zeta = 10^4$ (solid line) or $\zeta = 10^5$ (dashed line), and by a Weibull model (dotted curve), for the same data set used in the previous figure.

samples $Y^{*1}, \ldots, Y^{*M}$, where a bootstrap sample Y^{*k} is a random sample of size n drawn with replacement from the original sample Y. Then for the kth bootstrap sample, we compute an estimate of $S(t)$, call it $\hat{S}^*_k(t)$. Finally we compute the empirical standard deviation of the M samples and use it as an estimate for the standard error of $\hat{S}(t)$.

$$se_B = [\sum_{k=1}^{M}(\hat{S}^*_k(t) - \hat{S}^*_\cdot(t))^2/(M-1)]^{1/2},$$

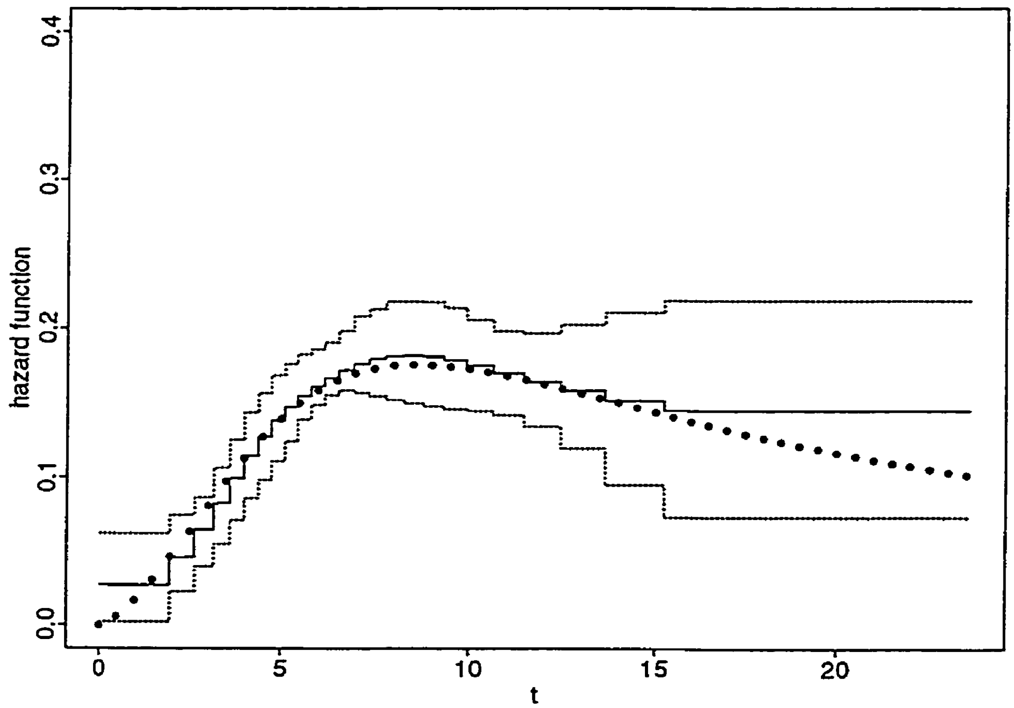

Figure 4.4: The dotted curve represents the true log-logistic hazard function used in simulations for Case 4. The solid line represents the average values of the 500 estimated hazard functions by a piecewise constant model. The dashed lines are the 0.025th and 0.975th percentiles of the 500 individual estimated hazard pieces by the piecewise constant model.

Table 4.2: Bias, asymptotic standard error (ASE) and sample standard error (SE) of estimates of survival function for T at five points based on 500 simulations, in Case 2: Weibull distribution, $\omega = 6$, $B = 17.99$

P-C model

t	S(t)	Bias ($\times 10^3$)	ASE ($\times 10^3$)	SE ($\times 10^3$)
1.8116	0.95	-13.8	40.3	27.1
4.7216	0.75	2.19	47.6	34.9
7.6960	0.5	1.61	52.8	33.0
11.311	0.25	-4.54	42.5	10.8
17.3548	0.05	3.76	27.0	12.8

where $\hat{S}^*_\cdot(t) = \frac{1}{M}\sum_{k=1}^{M} \hat{S}^*_k(t)$.

We generated five datasets of size 100 as in Case 1 of the simulation study in Section 4.3.1. For each dataset, we fitted a piecewise constant model with 25 pieces and the maximum penalized likelihood estimates of ρ were computed with tuning parameter $\zeta = 10^4$. $\hat{S}(t)$, where t is one of the 5th, 25th, 50th, 75th, 95th percentiles, was computed as in Section 4.3.1. Then the bootstrap estimates of the standard errors for $\hat{S}(t)$ were computed based on 100 bootstrap samples for each dataset. The results are listed in Table 4.5. We can see that the bootstrap estimates of the standard errors of $\hat{S}(t)$ are reasonably close to the sample standard errors of $\hat{S}(t)$ based on 500 samples in Table 4.1. This suggests that the bootstrap estimates of the standard errors are valid for this setting, and we recommend the use of bootstrap estimates of the standard errors when the maximum penalized likelihood approach is used. However, since we just computed a few examples here, more intensive study on the standard error estimates is required to support our

Table 4.3: Bias, asymptotic standard error (ASE) and sample standard error (SE) of estimates of survival function for T at five points based on 500 simulations, in Case 3: log-logistic distribution, $\omega = 1, B = 20.507$

P-C model

t	S(t)	Bias ($\times 10^3$)	ASE ($\times 10^3$)	SE ($\times 10^3$)
2.2665	0.95	0.187	23.6	20.3
4.7216	0.75	6.86	38.2	27.5
7.3080	0.5	2.81	46.7	19.4
11.311	0.25	0.446	40.3	7.91
23.5634	0.05	-8.34	24.3	10.7

recommendation.

Table 4.4: Bias, asymptotic standard error (ASE) and sample standard error (SE) of estimates of survival function for T at five points based on 500 simulations, in Case 4: log-logistic distribution, $\omega = 6, B = 20.507$

P-C model

t	S(t)	Bias ($\times 10^3$)	ASE ($\times 10^3$)	SE ($\times 10^3$)
2.2665	0.95	-14.4	44.0	32.6
4.7216	0.75	6.15	47.4	41.2
7.3080	0.5	0.413	52.1	26.8
11.311	0.25	-4.41	41.9	11.1
23.5634	0.05	-8.01	25.2	15.5

Table 4.5: Bootstrap estimates ($\times 10^3$) of the standard errors of $\hat{S}(t)$ based on piecewise constant hazard models for five datasets

t	$S(t)$	I	II	III	IV	V
1.8116	0.95	17.0	18.3	27.6	18.4	22.3
4.7216	0.75	21.4	27.7	31.6	32.2	37.8
7.6960	0.50	22.4	26.1	25.3	29.9	31.7
11.311	0.25	8.00	6.72	9.63	12.0	8.69
17.3548	0.05	9.37	6.18	7.12	10.6	7.79

Chapter 5

Conclusion and Further Research

5.1 Conclusion

In this thesis, we have discussed models with piecewise constant hazard, intensity or rate functions for event history data when event times are interval-censored, particularly for interval-grouped recurrent event data, current status data and doubly censored data.

We have presented two approaches for the interval-grouped recurrent event data. One is mixed Poisson process estimation; the other is a robust method that requires only specification of the mean structure and covariance structure among recurrent event counts. Both approaches incorporate piecewise constant rate functions. The robust method performs as well as the maximum likelihood method. However, variance estimation and confidence intervals based on maximum likelihood are non-robust, and the robust method should be used if the event process may not be a mixed Poisson process. The use of piecewise constant rate functions with just five to ten pieces provides excellent estimation of regression coefficients and mean functions.

We have developed methods that use piecewise constant hazard functions to estimate the distribution function of the induction time for standard current status data and doubly censored data. Regression models are also developed to assess covariate effects. Maximum likelihood estimates are obtained using a derivative-free optimization method. Our study shows that piecewise constant hazard models with five to ten pieces can provide good estimates of regression coefficients and survival functions when the time is not too small. These models are robust to the distributional form of the induction time. Therefore we recommend the use of the piecewise constant hazard models if we have any doubt about the distributional form of the induction time.

We have also investigated smoothing by the penalized likelihood approach. It is shown that combining smoothing and the piecewise constant models with more pieces can provide good estimates for the intensity, rate or hazard functions. Again this method is robust to the distributional form of the event time.

In summary, we recommend use of piecewise constant rate or hazard functions for event history data when event times are interval-censored and when we have doubt about the true form of rate functions or hazard functions. A model with five to ten pieces is usually good enough for the estimation of regression coefficients, mean functions or survival functions. For the estimation of rate functions or hazard functions, we recommend using of smoothing in a piecewise constant model with more pieces.

5.2 Further Research

There is further work to be done in this area. We outline some related problems in Section 5.2.1 and an extension of our method to zero-truncated recurrent event data in Section 5.2.2.

5.2.1 Some related problems

(1) Valid estimates of variation for maximum penalized likelihood estimates in the piecewise constant models.

In Chapter 4, we noticed that the standard errors for estimates derived from the minus second derivative matrices of the penalized log-likelihood may not be valid. Our solution to this problem is to use a bootstrap estimate of the standard error. However, it would be interesting to investigate the asymptotic theory of the maximum penalized likelihood estimates (MPLEs) in the piecewise constant models and derive an estimate of the covariance matrix of the MPLEs.

(2) Construction of confidence intervals for intensity, rate, or hazard functions when penalized likelihood approach is used.

This problem is closely related to the first problem. In Chapter 4, we constructed pointwise confidence intervals for hazard functions based on percentiles of estimated hazard function values in simulations. It would be useful to study whether we can construct confidence intervals based on asymptotic theory of the estimates and what are the coverage properties of these confidence intervals.

(3) Developing some diagnostic tools for model checking.

It is important to assess the model we have fitted. In Chapter 2, we have defined some residuals for the interval-censored recurrent event data. However, for the current status data and doubly censored data, it is difficult to define appropriate residuals, since lifetimes are never observed exactly for these data.

(4) An automatic procedure to choose the tuning constant in the penalized likelihood approach.

The tuning constant ζ in the penalized likelihood controls the level of smoothing. In our examples it was chosen by visually examining the smoothness of the estimates over some plausible values of ζ. An automatic procedure of selecting ζ would be useful. Marschner (1997) proposed choosing the tuning constants by minimizing a generalized cross-validation statistic Δ/k^2, where Δ is the (unpenalized) deviance corresponding to the penalized likelihood estimates and k is the degree of freedom as defined in Green (1987). However, this procedure requires the penalty matrices to be of a certain form, and under this condition, the maximization of the penalized likelihood is equivalent to fitting a model where the parameters are cubic spline functions. Joly *et al.* (1998) selected the tuning constants by maximizing an approximate cross-validation score based on log-likelihood. Their approach requires approximating the solution of the maximum of the penalized likelihood on a basis of splines. So none of their formulations apply directly to our situation, since we do not want to assume ρ_k to be a spline function.

5.2.2 Zero-truncated recurrent event data

Hu and Lawless (1996) have considered estimation from zero-truncated recurrent event data where the event process $\{N_i(t), t > 0\}$ of the ith unit has an observation window $(0, \tau_i]$ and the process is unknown to the observer unless at least one event has occurred in $(0, \tau_i]$. The total number of units under consideration is M but the value of M is unknown. The available data include the time of events t_{ij} $(j = 2, \ldots, n_i)$ and end of observation time τ_i for unit i, $i = 1, \ldots, m$. Hu and Lawless (1996) have proposed several approaches to estimate the rate and mean functions, including zero-truncated Poisson process models with piecewise constant rate assumption. Below we outline application of mixed Poisson process models to the problem.

We assume that $\{N_i(t), t > 0\}$, $i = 1, \ldots, m$, are independent Poisson processes with intensity function $\alpha_i \lambda_0(t)$, wherer $\alpha_i, i = 1, \ldots, n$ are independent Gamma variables with mean 1 and variance v. Let $G(x)$ be the CDF of α and $\Lambda(t) = \int_0^t \lambda(s) ds$ be the cumulative intensity function. $\Lambda(t)$ is also the mean function of $\{N_i(t)\}$. The likelihood based on observed data $t_{ij}, j = 1, \ldots, n_i (n_i > 0)$ and τ_i, $i = 1, \ldots, m$ is

$$\begin{aligned} L &= \prod_{i=1}^{m} L_i = Prob\{t_{i1}, \ldots, t_{in_i} | n_i > 0, \tau_i\} \\ &= \prod_{i=1}^{m} \frac{\int_0^\infty \prod_{j=1}^{n_i} \lambda_0(t_{ij}) \alpha_i^{n_i} \exp(-\alpha_i \Lambda_0(\tau_i)) dG(\alpha_i)}{\int_0^\infty [1 - \exp(-\alpha_i \Lambda_0(\tau_i))] dG(\alpha_i)}. \end{aligned} \tag{5.1}$$

It can be simplified as

$$L = \prod_{i=1}^{m} \frac{\prod_{j=1}^{n_i} \lambda_0(t_{ij}) v^{n_i} \prod_{k=1}^{n_i} (v^{-1} + k - 1)}{((1 + v\Lambda_0(\tau_i))^{1/v} - 1)(1 + v\Lambda_0(\tau_i))^{n_i}}. \tag{5.2}$$

Suppose a piecewise constant intensity function is used, that is, $\lambda_0(t) = \rho_k$, if $t \in (a_k, a_{k+1}]$, for $0 = a_1 < a_2 < \ldots < a_r$. Then $\Lambda_0(t) = \sum_{k=1}^{r} \rho_k u_k(t)$, where

$u_k(t) = max(0, min(a_{k+1}, t) - a_k))$. The log-likelihood function is

$$\begin{aligned} l &= \sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{k=1}^{r} \delta_k(t_{ij}) \log \rho_k + \sum_{i=1}^{m} \log v \\ &+ \sum_{i=1}^{m}\sum_{k=1}^{n_i} \log(v^{-1} + k - 1) - \sum_{i=1}^{m} n_i \log(1 + v\Lambda_0(\tau_i)) \\ &- \sum_{i=1}^{m} \log((1 + v\Lambda_0(\tau_i))^{1/v} - 1), \end{aligned} \tag{5.3}$$

where $\delta_k(t) = 1$ if $t \in (a_k, a_{k+1}]$; $\delta_k(t) = 0$ otherwise. The log-likelihood function can be maximized by a derivative-free optimization algorithm to get the maximum likelihood estimates of $(\rho_1, \ldots, \rho_r, v)$. The estimate of the mean function can be computed as $\hat{\Lambda}_0(t) = \sum_{k=1}^{r} \hat{\rho}_k u_k(t)$. Inference can made based on the asymptotic normality of the maximum likelihood estimates.

Bibliography

Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, **6**, 701–726.

Aalen, O. O. (1994). Effects of frailty in survival analysis. *Statistical Methods in Medical Research*, **3**, 227–243.

Aalen, O. O. and Husebye, E. (1991). Statistical analysis of repeated events forming renewal processes. *Statistics in Medicine*, **10**, 1227–1240.

Abu-Libdeh, H., Turnbull, B. W., and Clark, L. C. (1990). Analysis of multi-type recurrent events in longitudinal studies; application to a skin cancer prevention trial. *Biometrics*, **46**, 1017–1034.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki, editors, *2nd International Symposium on Information Theory*. Akademiai Kiado, Budapest.

Andersen, P. K. *et al.* (1993). *Statistical models based on counting processes.* Springer-Verlag, New York.

Bacchetti, P. (1990). Estimating the incubation period of AIDS by comparing population infection and diagnosis patterns. *Journal of the American Statistical Association*, **85**, 1002–1008.

Bacchetti, P. and Jewell, N. P. (1991). Nonparametric estimation of the incubation period of AIDS based on a prevalent cohort with unknown infection times. *Biometrics*, **47**, 947–960.

Breslow, N. E. (1990). Tests of hypotheses in overdispersed Poisson regression and other quasi-likelihood models. *Journal of the American Statistical Association*, **85**, 565–571.

Byar, D. P. (1980). The veterans administration study of chemoprophylaxis for recurrent stage I bladder tumors: comparisons of placebo, pyridoxine, and topical thiotepa. In M. Pavone-Macaluso, P. H. Smith, and F. Edsmyr, editors, *Bladder tumors and other topics in urological oncology*, pages 363–370. Plenum, New York.

Carstensen, B. (1996). Regression models for interval censored survival data: Application to HIV infection in danish homosexual men. *Statistics in Medicine*, **15**, 2177–2189.

Clayton, D. (1994). Some approaches to the analysis of recurrent event data. *Statistical Methods in Medical Research*, **3**, 244–262.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74**, 829–836.

Coates, R. A., Farewell, V. T., *et al.* (1990). Cofactors of progression to AIDS in a cohort of male sexual contacts with men with HIV disease. *American Journal of Epidemiology*, **132**, 717–722.

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

Cox, D. R. (1972b). The statistical analysis of dependencies in point processes. In P. A. W. Lewis, editor, *Symposium on Point Processes.* wiley, New York.

Cox, D. R. and Oakes, D. (1984). *Analysis of survival data.* Chapman and Hall, London.

De Gruttola, V. G. and Lagakos, S. W. (1989). Analysis of doubly-censored survival data, with application to AIDS. *Biometrics*, **45**, 1–11.

Dean, C. B. (1991). Estimating equations for mixed Poisson models. In V. P. Godambe, editor, *Estimating Functions*, pages 35–46. Clarendon Press, Oxford.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B*, **39**, 1–38.

Diamond, I. D. and Mcdonald, J. W. (1992). Analysis of current-status data. In J. Trussell *et al.*, editors, *Demographic application of event history analysis*, pages 231–252. Clarendon Press, Oxford.

Efron, B. and Tibshirani, R. J. (1993). *An introduction to the bootstrap.* Chapman & Hall, New York.

Fleming, T. R. and Harrington, D. R. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

Follmann, D. A. and Goldberg, M. S. (1988). Distinguishing heterogeneity from decreasing hazard rates. *Technometrics*, **30**, 389–396.

Frydman, H. (1995). Semiparametric estimation in a three-state duration-dependent markov model from interval-censored observations, with application to AIDS data. *Biometrics*, **51**, 502–511.

Fusaro, R. E., Bacchetti, P., and Jewell, N. P. (1996). A competing risks analysis of presenting AIDS diagnoses trends. *Biometrics*, **52**, 211–225.

Green, P. J. (1987). Penalized likelihood for generalized semi-parametric regression models. *International Statistical Review*, **55**, 245–259.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach*. Chapman and Hall, London.

Gruger, J., Kay, R., and Schumacher, M. (1991). The validity of inferences based on incomplete observations in disease state models. *Biometrics*, **47**, 595–605.

Heilbron, D. C. (1989). Generalized linear models for altered zero probabilities and overdispersion in count data. Technical report, University of California, San Francisco, Department of Epidemiology and Biostatistics.

Hu, X. J. and Lawless, J. F. (1996). Estimation of rate and mean functions from truncated recurrent event data. *Journal of the American Statistical Association*, **91**, 300–310.

Jewell, N. P. and van der Laan, M. (1996). Generalizations of current status data with applications. In N. P. J. et al., editor, *Lifetime Data: Models in Reliability and Survival Analysis*, pages 141–148. Kluwer Academic Publishers, Netherlands.

Jewell, N. P., Malani, H. M., and Vittinghoff, E. (1994). Nonparametric estimation for a form of doubly censored data, with application to two problems in AIDS. *J. of the American Statist. Association*, **89**, 7–18.

Joly, P., Commenges, D., and Letenneur, L. (1998). A penalized likelihood approach for arbitrarily censored and truncated data: application to age-specific incidence of dementia. *Biometrics*, **54**, 185–194.

Kalbfleisch, J. D. *et al.* (1991). Methods for the analysis and prediction of warranty claims. *Technometrics*, **33**, 273–285.

Kalbfleisch, J. D. and Lawless, J. F. (1989). Inference based on retrospective ascertainment: an analysis of the data on transfusion-related AIDS. *Journal of the American Statistical Association*, **84**, 360–372.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *Statistical analysis of failure time data*. Wiley, New York.

Kaufmann, H. (1987). Regression model for nonstationary categorical time series; asymptotic estimation theory. *The Annals of Statistics*, **15**, 79–98.

Kim, D. K. (1997). Regression analysis of interval-censored survival data with covariates using log-linear models. *Biometrics*, **53**, 1274–1283.

Kim, M. Y., De Gruttola, V. G., and Lagakos, S. W. (1993). Analyzing doubly censored data with covariates, with application to AIDS. *Biometrics*, **49**, 13–22.

Klein, J. P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics*, **48**, 795–806.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, **34**, 1–14.

Lawless, J. F. (1982). *Statistical models and methods for lifetime data.* Wiley, New York.

Lawless, J. F. (1987a). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics*, **15**, 209–225.

Lawless, J. F. (1987b). Regression methods for Poisson process data. *Journal of the American Statistical Association*, **82**, 808–815.

Lawless, J. F. (1995). The analysis of recurrent event for multiple subjects. *Appl. Statist.*, **44**, 487–498.

Lawless, J. F. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics*, **37**, 158–168.

Liang, K.-Y. *et al.* (1993). Modelling marginal hazards in multivariate failure time data. *J. R. Statist. Soc. B*, **55**, 441–454.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.

Lin, D. Y. (1994). Cox regression analysis of multivariate failure time data: the marginal approach. *Statistics in Medicine*, **13**, 2233–2247.

Lindsey, J. C. and Ryan, L. M. (1993). A three-state multiplicative model for rodent tumorigenicity experiments. *Applied Statistics*, **42**, 283–300.

Marschner, I. C. (1997). A method for assessing age-time disease incidence using serial prevalence data. *Biometrics*, **53**, 1384–1398.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall.

Ng, E. T. M. and Cook, R. J. (1997). Adjusted score tests of homogeneity for Poisson processes. Technical report, University of Waterloo, Department of Statistics and Actuarial Science.

O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statistical Science*, **1**, 502–527.

Pickles, A. and Crouchley, R. (1994). Generalizations and applications of frailty models for survival and event data. *Statistical Methods in Medical Research*, **3**, 263–278.

Press, W. H., Flannery, B. P., *et al.* (1990). *Numerical Recipes*. Cambridge, Cambridge.

Rabinowitz, D., Tsiatis, A., and Aragon, J. (1995). Regression with interval-censored data. *Biometrika*, **82**, 501–513.

Rosenberg, P. S. (1995). Hazard function estimation using b-splines. *Biometrics*, **51**, 874–887.

Ross, S. (1983). *Stochastic Processes*. Wiley, New York.

Schluchter, M. D. and Jackson, K. L. (1989). Log-linear analysis of censored survival data with partially observed covariates. *Journal of the American Statistical Association*, **84**, 42–52.

Segal, M. R., Bacchetti, P., and Jewell, N. P. (1994). Variances for maximum penalized likelihood estimates obtained via the EM algorithm. *J. R. Statist. Soc. B*, **56**, 345–352.

Severini, T. A. and Wong, W. H. (1992). Profile likelihood and conditionally parametric models. *Annals of Statistics*, **20**, 1768–1802.

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

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *J. R. Statist. Soc. B*, **47**, 1–52.

Staniswalis, J. G., Thall, P. F., and Salch, J. (1997). Semiparametric regression analysis for recurrent event interval counts. *Biometrics*, **53**, 1334–1353.

Sun, J. (1995). Empirical estimation of a distribution function with truncated and doubly interval-censored data and its application to AIDS studies. *Biometrics*, **51**, 1096–1104.

Sun, J. and Kalbfleisch, J. D. (1993). The analysis of current status data on point processes. *Journal of the American Statistical Association*, **88**, 1449–1454.

Sun, J. and Kalbfleisch, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Statistica Sinica*, **5**, 279–290.

Sun, J. and Matthews, D. E. (1997). Random-effect regression model for medical follow-up studies. *Canadian Journal of Statistics*, **25**, 101–111.

Sun, J. and Wei, L. J. (1996). Regression analysis for panel count data. Technical report, Harvard University, Department of Biostatistics.

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

Thall, P. F. and Lachin, J. M. (1988). Analysis of recurrent events: Nonparametric methods for random-interval count data. *Journal of the American Statistical Association*, **83**, 339–347.

Thall, P. F. and Vail, S. C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics*, **46**, 657–671.

Ware, J. M. *et al.* (1988). Issues in the analysis of repeated categorical outcomes. *Statistics in Medicine*, **7**, 95–107.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*, **84**, 1065–1073.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–25.

Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics*, **11**, 95–103.

Yan, P. and Lawless, J. F. (1992). Non-parametric estimation of a survival distribution when time of onset is unknown. Technical report, Department of Statistics and Actuarial Science, University of Waterloo.