ON COMPETING RISKS DATA WITH COVARIATES AND LONG-TERM SURVIVORS

TAN ZHIPING

Ph.D

The Hong Kong Polytechnic University

2011

The Hong Kong Polytechnic University

Department of Applied Mathematics

On Competing Risks Data with Covariates and Long-term Survivors

TAN ZHIPING

A thesis submitted in partial fulfilment of
the requirements for the degree of Doctor of Philosophy

Mar. 2010

CERTIFICATE OF ORIGINALITY

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of a university or other institute of higher learning, except where due acknowledgment is made in the text.

TAN ZHIPING

Abstract

In the study of survival data, one of the important problems is about competing risks combined with the possible existence of long-term survivors (subjects that will never experience the events under consideration, also referred to as "cured" or "immunes"). Under this scenario, it is important to know the failure rates with respect to different risks and the cured proportion. It is also useful to make inferences on the regression coefficients of the covariates that influence the failure. Further analysis of the significance levels of the parameters plays an important role in the study. In particular, a major attention is paid to the significance level of the cured rate which implies the existence of immunes.

In this dissertation, three models are investigated for survival data with competing risks, covariates and immunes: general mixture model, piecewise exponential mixture model and proportional cause-specific hazards model.

In the general mixture setting, full maximum likelihood methods are employed to draw statistical inferences on the model attributes and the asymptotic properties of the estimators. Likelihood ratio tests are developed to test the significance levels of

the parameters and the relationships among them. Under some regularity conditions and mild assumptions, the estimators are proved to be consistent and asymptotically normally distributed, and the tests are also consistent and follow different distributions according to the underlying hypotheses. The performances of the estimators and tests are assessed by a simulation study. It shows that the approach given in this part provides a satisfactory way to investigate many practical problems.

The second part of the dissertation is the piecewise exponential mixture model for competing risks data. The existence, consistency and asymptotic normality of the estimators are rigorously derived under general sufficient conditions. Likelihood ratio tests are investigated for various hypotheses of practical interest. A study of real life data is conducted to illustrate the approach.

In addition, a semi-parametric approach is proposed to investigate the competing risks data under the assumptions of independent censoring and proportional cause-specific hazard functions among different risks and covariates. Partial likelihood methods are used to make inferences on the levels of the risks.

Acknowledgements

First and foremost, my deepest gratitude goes to my chief supervisor, Dr. Wai-cheung Ip, for the guidance and encouragement he gave me in my studies. Without his constant and illuminating instructions, this thesis could not have reached its present form. During the period of my studies, he not only taught me how to overcome difficulties in my research, but also helped me to understand what a person of exemplary virtue is. Second, I would like to express my heartfelt gratitude to Dr. Xian Zhou, my external co-supervisor, for his valuable advice and warm care in the past years. It is also my great pleasure to thank Dr. Xun Li, my co-supervisor, for his insightful comments and helpful advice in my study. My biggest thanks must go to my family for their constant love and the great confidence that they have shown in me throughout these years. Last but not least, I would like to thank the Department of Applied Mathematics for its excellent facility and friendly staff and students. Because of them, it has been a precious experience for me to study for my PhD degree in the Department of Applied Mathematics at The Hong Kong Polytechnic University.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The focus of survival analysis is the study of the time duration until the occurrence of a certain event of interest, such as death of a patient or a life insurance policyholder, recurrence or relapse of a disease, lodgment of an insurance claim, a natural disaster, settlement of a liability law suit, malfunction of a consumer product, a failure of a company, rearrest of a former prisoner, marriage breakdown, and so on.

The occurrence of the event of interest is often referred to as a *failure* (although it may actually represent a positive outcome in some applications), and the time duration until it occurs is called the *failure time*. Failure time, however, is not the only thing of concern. Also of great interest in survival analysis are *failure causes*, referred to as *competing risks* for there are often different types of risks that are "competing" to cause the failure. Take an employee benefit plan for example, the present value of the benefits to be paid depends on both the time and the cause of the termination of employment (retirement, resignation, death, disability, etc.). Other examples include different causes of death, different types of losses under insurance coverage, and so on.

Another important issue in survival analysis is the possible presence of *long-term survivors* (also known as *immunes* or *cured* individuals) in the sense that they are not subject to the event of interest. This is natural in many real life situations, such as patients cured of a life-threatening disease, or insured employees who never experience disability. In such situations, not only the failure mechanism with respect to the risks but also the proportion of long-term survivors (such as the cured rate of the patients under a certain treatment) is of interest.

Furthermore, the failure time and failure cause of an individual are often influenced by certain particular features associated with the individual, which are referred to as *covariates*. In medical study, for example, covariate factors are often comprised of age, gender, health condition, treatment method, etc. for a patient under observation. In motor vehicle insurance, age and postal code are often taken as major indicators in assessing the risk level of a driver. For the application of the models in real life, it is important to consider covariate information as it usually exists and influences the outcome of analysis.

Owing to the wide application of survival analysis, there have been many investigations of survival data with competing risks or immunes in the presence of covariates. However, the large sample analysis of survival data, which incorporate competing risks, covariates and possible presence of immunes, has been limited to parametric models. In addition, the cases where the "true" value of parameters may be on the boundary of the parameter space, (e.g. the cured rate is not significant) have not been studied together with covariates. In this study, we will attempt to narrow this gap with extensive and thorough investigations to general mixture model and proportional cause-specific hazards model for life time data mentioned above. Questions of interest are:

1. What is the probability of a subject to fail from a certain risk?
2. What is the cumulative distribution function of an individual conditional on a given risk?
3. Is there any insignificant covariate factor that can be dropped from the analysis?
4. Are the baseline cause-specific hazard rates proportional to each other so that a proportional hazards model is suitable?
5. Do immunes exist? If yes, what is their proportion in the population?

The remainder of this thesis is devoted to answer these five questions. In Chapter 2, the previous work of competing risks and immunes is reviewed and the differences between our work and previous work are discussed. In Chapter 3, a series of general mixture models are reported with answers developed for these five questions. It is proved that the mixture model setting produces consistent and asymptotically normally distributed estimators under some mild sufficient conditions when the true values of the parameters lie on the parameter space (interior or boundary) and questions 1 and 2 are answered consequently. For questions 3-4, the large sample properties of the likelihood ratio statistics are investigated to test the significance levels of the parameters and the relationships among them. The consistency of the tests are proved and their large sample distributions are provided. In particular, the asymptotic distribution of the likelihood ratio statistic for the boundary hypothesis of the exhaustiveness of failure causes (no cured individuals or immunes) is investigated. Question 5 is answered by the result of the test.

An application of the general results on real life data is demonstrated in Chapter 4. A piecewise exponential mixture model for competing risks data is proposed, which allows censoring, covariates and immunes. The maximum likelihood estimators are shown to be unique local to the true values of the parameters with probability approaching 1, consistent and asymptotically normal. To assess the performances of our models, a set of Bone Marrow Transplant Data is fitted by the proposed model.

In Chapter 5, we propose a semi-parametric approach to draw statistical inferences, under the assumption that the cause-specific failure rates are proportional to each other. The estimators are proved to be consistent, and the estimated coefficients of covariates have asymptotic normal distributions whereas the estimated coefficients of risks have asymptotic log-normal distributions in the interior of the parameter space. A simulation is conducted to examine the proposed methods.

The main contributions of the thesis are summarized as follows:

(i) A general class of mixture models are developed, which account for competing risks, possible presence of long-term survivors and covariates.

(ii) Maximum likelihood estimators of the model parameters as well as deviance tests for the presence of long-term survivors and covariate effects are established, and their large-sample properties are proved rigorously.

(iii) A class of piecewise exponential mixture models is investigated with rigorous derivation of statistical inference and application to real-life data.

(iv) A semi-parametric proportional hazards model is developed and the method of partial likelihood is employed to draw statistical inference on competing risks, proportions of long-term survivors and covariate effects.

Chapter 2

Literature Review

The origin of the study on competing risks can be traced back to Daniel Bernoulli's (1760) attempt to separate the cause of death by smallpox from other causes. Since then, research in this area has been developed quickly due to its wide applications in many areas such as medical research, health science, actuarial science, economics, finance, management, engineering reliability, criminology, social science, etc. The study of competing risks has now become one of the issues of primary importance in the analysis of survival data.

In the competing risks model, individuals are exposed to several distinct causes of failures, but if a failure happens, it can only be due to one of the risks of interest. Cox (1959) formulated the survival data with competing risks in terms of latent failure times. The risks are assumed to be independent of each other. Although there are various criticisms on the identifiability problems of this assumption, models with latent variables have been commonly used in data analyses since then, especially in the areas of economics and finance. Literature on latent variables applied to various fields can be found, for example, in Han and Hausman (1990) and references therein. An alternative approach for survival data with competing risks is the cause-specific hazard rates setting which is proposed by Prentice *et al* (1978). This framework allows for dependence between competing risks. The survival analysis based on the hazard functions has been popular, as they can approximate survival data more naturally than the latent failure times setting. One of the most widely used models is the Cox's proportional hazards approach of Cox (1972), in which the hazard function is assumed to be

$$\lambda(t|Z) = \lambda_0(t) \exp(\beta^T Z),$$

where λ_0 is a baseline hazard function, β is a vector of regression parameters and $\lambda(t|Z)$ is the hazard function for an individual with a κ-dimensional covariate vector Z.

The concept of immunes or cured individuals has also been investigated for several decades. Boag (1949) is among the first to study the cured proportion in cancer patients. Since the cured rate cannot be estimated independently from the other parameters of the model in the presence of censoring, a commonly used approach to solve this problem is to formulate the model as a mixture of two populations, one for cured individuals or immunes and the other for "susceptibles" (individuals suffering from the risks of interest). An extensive discussion of popular methods of inference on this model has been provided by Maller and Zhou (1996). Such mixture models have also been investigated by many other authors. Among them Peng and Dear (2000) adopted a non-parametric approach, Ghitany and Maller (1992) and Ghitany et al. (1994) studied exponential mixture distributions and Vu et al. (1998) extended their work to survival distributions mixed by exponential family and cured rate of the population. In addition, semi-parametric mixture models are discussed by Zhao and Zhou (2006).

Various models of different types have been proposed for survival data with immunes and competing risks. Miyakawa (1984) gave parametric estimation for two failure causes and allow possible missing causes, Kundu and Basu (2000) studied their large sample properties. Larson and Dinse (1985) suggested a mixture model approach and computed maximum likelihood estimators by replacing the right-censored failure times under unknown causes with a set of suitably weighted hypothetical ones under known causes. Maller and Zhou (2002) provided a rigorous analysis of the parametric mixture models and derived useful large sample properties (consistency and asymptotic normality) of the maximum likelihood estimators and the asymptotic distributions for the test statistics under certain boundary hypotheses. However their work was restricted to the i.i.d. case without covariates. As a complementarity, Choi and Zhou (2002) studied the large sample properties of a class of parametric mixture models with covariates for competing risks in the presence of immunes.

There are a number of similarities among the works of Larson and Dinse (1985), Maller and Zhou (2002), Choi and Zhou (2002) and the present work. The current approach extends the works of the above authors by relaxing their restrictions in several ways.

Suppose that individuals are exposed to J causes and the failure time T of an individual with covariate Z has a mixture distribution as follows:

$$F(t|Z) = P\{T \leq t|Z\} = \sum_{j=1}^{J} p_j(Z)F_j(t|Z)$$

with $\sum_{j=1}^{J} p_j(Z) \leq 1$, where $p_j(Z)$ is the probability that this individual fails from risk j and $F_j(t|Z)$ is the failure distribution function for this individual, conditional on the eventual failure from risk j. The approach proposed by Larson and Dinse (1985) is restricted to the boundary case in the sense that $\sum_{j=1}^{J} p_j(Z) = 1$, i.e. it does not allow for immunes. The models of Maller and Zhou (2002) do allow for the existence of immuned individuals by a possibly improper setting but not covariates. Finally, Choi and Zhou (2002) attempted an improper mixture model with $\sum_{j=1}^{J} p_j(Z) < 1$ for competing risks data with covariates and immunes included. Their work, however, assumes the existence of immunes as a priori. This yields identifiability problems as it is hard to identify such kind of information from the observations. For instance, if the censoring is heavy, we can only consider the probable existence of immunes but cannot claim that there must be immunes as censoring is also influenced by experimental design and other reasons.

It is desirable and of practical interest to relax the restrictions of the above works so that a more general approach can better model real life data. In this dissertation, a general mixture model is developed to cover the models of Larson and Dinse (1985), Maller and Zhou (2002) and Choi and Zhou (2002), with extensive and thorough investigations of the large sample properties of the estimators and various test statistics. The model is more flexible than the previous models and can be readily applied to real life problems. For the case when the data have good properties of proportional hazards, a semi-parametric approach is investigated, which is relatively more restrictive than the general mixture model but much easier to use.

Chapter 3

General Mixture models

3.1 Introduction

In this chapter, a general mixture model is introduced to make inferences on competing risks data with covariates and possible immunes. In Section 3.2, the framework of the model is explained. In Section 3.3, the large sample properties of the estimators and test statistics are provided with the explanation of the applications. A simulation is shown in Section 3.4 to assess the performance of the methods. Section 3.5 is devoted to the discussions and future work. Finally, the proofs are collected in Section 3.6.

3.2 Model Specification

Suppose there are n individuals who are exposed to J distinct failure causes and the observation of an individual can be either censored or caused by one and only one of these risks. Further assume that there may be individuals who are not susceptible to the risks under consideration, i.e. immunes may exist. The competing risks data in the presence of censoring consist of observations $(t_i, \delta_{ij}, z_i), i = 1, \ldots, n$, where t_i is the observed failure time of individual i, z_i is the covariate information for individual i, which is also observable and for $i = 1, \ldots, n$ and $j = 1, \ldots, J$,

$$\delta_{ij} = \begin{cases} 1 & \text{if individual } i \text{ dies from cause } j, \\ 0 & \text{otherwise} \end{cases}$$

are the indicators of the failure causes. For each individual i, the indicator of the censoring status is defined as

$$\delta_i = \begin{cases} 1 & \text{if individual } i \text{ is uncensored,} \\ 0 & \text{if individual } i \text{ is censored.} \end{cases}$$

Thus

$$\delta_i = \sum_{j=1}^{J} \delta_{ij}, \quad i = 1, 2, \ldots, n.$$

Further suppose that the individuals under study are independent of each other and

$$p_{ij} = P\{\text{individual } i \text{ will fail from cause } j \text{ eventually}\}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, J$. Thus

$$p_i = \sum_{j=1}^{J} p_{ij} = P\{\text{ individual } i \text{ is susceptible to the risks under consideration}\},$$

where $0 < p_i \leq 1$ and the existence of cured individuals or immunes is implied by $p_i < 1$.

For $i = 1, \ldots, n$, let T_i^* be a nonnegative independent random variable representing the true failure time (without censoring) of individual i, u_i be the censoring time of individual i and t_i^* be the realization of T_i^*. It is obvious that

$$t_i = u_i \wedge t_i^*.$$

Further associate each individual i a random variable D_i with

$$D_i = \begin{cases} j & \text{if individual } i \text{ will fail from cause } j \text{ eventually,} \\ 0 & \text{individual } i \text{ is an immune subject} \end{cases}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, J$. Here we can treat D_i as a discrete random variable with the following proper distribution function:

$$P\{D_i = j\} = p_{ij},\ 1 \leq j \leq J \qquad \text{and} \qquad P\{D_i = 0\} = 1 - p_i.$$

Thus D_i can be interpreted as an indicator of the failure cause of individual i. It should be pointed out that the value of D_i is observed when $\delta_i = 1$ and cannot be observed when $\delta_i = 0$. Throughout our study, an i.i.d. censoring mechanism is assumed, i.e. u_i's are independent of each other and they are all independent of the t_i^*'s. Thus

$$\delta_{ij} = 1_{\{t_i^* \leq u_i, D_i = j\}} \qquad \text{and} \qquad \delta_i = 1_{\{t_i^* \leq u_i\}}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, J$, where 1_E is the indicator of event E which takes value 1 when E occurs and 0 otherwise. And the c.d.f. of the censoring time is assumed to be non-informative. Let

$$F_{ij}(t) = P\{T_i^* \leq t | D_i = j\} \tag{3.2.1}$$

represent the conditional c.d.f. of T_i^* conditional on the type j failure for $i = 1, \ldots, n$ and $j = 1, \ldots, J$. It is obvious that $F_{ij}(t)$ is proper. Also let

$$F_i(t) = P\{T_i^* \leq t\} \tag{3.2.2}$$

denote the unconditional c.d.f. of T_i^*. A fact to be noted is that

$$P\{T_i^* = \infty | D_i = 0\} = 1. \tag{3.2.3}$$

Then $F_i(t)$ can be given by

$$F_i(t) = \sum_{j=0}^{J} P\{T_i^* \leq t | D_i = j\} P\{D_i = j\} = \sum_{j=1}^{J} p_{ij} F_{ij}(t). \tag{3.2.4}$$

When $p_i = \sum_{j=1}^{J} p_{ij} < 1$, $F_i(t)$ is improper. We shall assume that $F_{ij}(t)$ and $F_i(t)$ have density functions $f_{ij}(t)$ and $f_i(t)$ respectively throughout, this is not essential but it simplifies the notations.

The above specifications provide a probabilistic foundation for a general mixture model approach to accommodate competing risks data with possible immunes. Like Maller and Zhou (2002), we further provide a cause-specific formulation approach which is equivalent to the general mixture approach and also widely used in survival analysis. For the details of the cause-specific formulations, see Prentice *et al.* (1978) and Kalbfleisch and Prentice (1980, p. 167). The cause-specific hazard function associated with individual i and risk j is defined by

$$\lambda_{ij}(t)dt = P\{T_i^* \in [t, t + dt), D_i = j | T_i^* \geq t\}. \tag{3.2.5}$$

The hazard function for individual i is

$$\lambda_i(t)dt = \sum_{j=1}^{J} \lambda_{ij} dt = P\{T_i^* \in [t, t + dt) | T_i^* \geq t\}. \tag{3.2.6}$$

The hazard functions can be expressed in terms of p_{ij}, $f_{ij}(t)$ and $F_i(t)$ as

$$\lambda_{ij}(t) = \frac{p_{ij} f_{ij}(t)}{1 - F_i(t)} \tag{3.2.7}$$

and

$$\lambda_i(t) = \frac{f_i(t)}{1 - F_i(t)} \tag{3.2.8}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, J$. Conversely, p_{ij} and $F_{ij}(t)$ can be defined in terms of λ_{ij} and $F_i(t)$ as:

$$p_{ij} = P\{D_i = j\} = \int_0^\infty \lambda_{ij}(t)(1 - F_i(t))dt \tag{3.2.9}$$

and

$$F_{ij}(t) = \frac{P\{T_i^* \leq t, D_i = j\}}{P\{D_i = j\}} = \frac{1}{p_{ij}} \int_0^t \lambda_{ij}(y)(1 - F_i(y))dy. \tag{3.2.10}$$

Thus for a fixed $i \in \{1, \ldots, n\}$ and $1 \leq j \leq J$, $\lambda_{ij}(t)$'s are proportional to each other if and only if $f_{ij}(t)$'s are proportional to each other.

Apart from a multiplicative constant, the full likelihood of the n observations is given by

$$\begin{aligned} L_f &= \prod_{i=1}^n \{ \left(p_{ij(i)} f_{ij(i)}(t_i)\right)^{\delta_i} \left(1 - F_i(t_i)\right)^{1-\delta_i} \} \\ &= \prod_{i=1}^n \left\{ \prod_{j=1}^J \left(p_{ij} f_{ij}(t_i)\right)^{\delta_{ij}} \left(1 - F_i(t_i)\right)^{1-\delta_i} \right\}, \end{aligned} \tag{3.2.11}$$

where $j(i)$ indicates the observed failure cause of individual i when a failure occurs. And the log-likelihood function is

$$l_n = \sum_{i=1}^n \left\{ \sum_{j=1}^J \delta_{ij} \log\left(p_{ij} f_{ij}(t_i)\right) + (1 - \delta_i) \log\left(1 - F_i(t_i)\right) \right\}. \tag{3.2.12}$$

Thus individual i contributes $p_{ij} f_{ij}(t_i)$ to the likelihood function if he/she experiences a type j failure. Alternatively, the contribution to the likelihood function of individual i with censored failure time observed should be $S_i(t_i) = 1 - F_i(t_i)$.

Assume that

$$f_{ij}(t) = f(t; \phi_{ij}), \qquad t \geq 0, \tag{3.2.13}$$

where $\phi_{ij} = [\phi_{ij1,}\phi_{ij2}, \ldots, \phi_{ijK}]^T$ is a K-vector of parameters varying over an open subset of $\mathbb{R}^K$, and T denotes the transpose of a matrix or vector. Suppose that for $1 \leq i \leq n$, $1 \leq j \leq J$ and $1 \leq k \leq K$, ϕ_{ijk}'s are linked to the linear combinations of covariates by

$$\phi_{ijk} = \eta(\kappa_{ijk}) \quad \text{with} \quad \kappa_{ijk} = \alpha_{jk}^T x_i, \tag{3.2.14}$$

where x_i is a π_1-dimensional sub-vector of z_i, α_{jk} is a π_1-vector of parameters and η is a differentiable function from $\mathbb{R}$ to the domain of ϕ_{ijk}. Let

$$p_{ij} = \frac{\zeta(\xi_j + \rho_{ij})}{\omega(\varrho_i) + \sum_{l=1}^{J} \zeta(\xi_l + \rho_{il})}, \quad \text{with} \qquad \rho_{ij} = \beta_j^T y_i \quad \text{and} \quad \varrho_i = \gamma^T w_i,$$

where ζ is a link functions from $\mathbb{R}$ to $\mathbb{R}^+$ and ω is a real function. As above, y_i and w_i are π_2-dimensional and π_3-dimensional sub-vectors of z_i respectively with corresponding regression coefficients β_j and γ. Thus α_{jk}, ξ_j, β_j and γ, $1 \leq j \leq J$, $1 \leq k \leq K$ constitute parameters to be estimated from the data. Let θ be a vector that consists of all parameters, thus

$$\theta = (\alpha_1^T, \ldots, \alpha_J^T, \beta_1^T, \ldots, \beta_J^T, \xi_1, \ldots, \xi_J, \gamma)^T \in \mathbb{R}^{\pi_1 JK + (\pi_2+1)J + \pi_3},$$

where $\alpha_j^T = (\alpha_{j1}^T, \ldots, \alpha_{jK}^T) \in \mathbb{R}^{\pi_1 K}$, $\beta_j^T \in \mathbb{R}^{\pi_2}$, $\xi_j \in \mathbb{R}$ and $\gamma^T \in \mathbb{R}^{\pi_3}$. The true value of the parameter vector θ is

$$\theta_0 = (\alpha_{10}^T, \ldots, \alpha_{J0}^T, \beta_{10}^T, \ldots, \beta_{J0}^T, \xi_{10}^T, \ldots, \xi_{J0}^T, \gamma_0^T)$$

with γ_0 constrained in

$$\Gamma = \{\gamma \,|\, \gamma \in \mathbb{R}^{\pi_3}, \, \omega(\gamma^T w_i) \geq 0, \text{ for all } 1 \leq i \leq n\}.$$

Thus the parameter space Θ for θ is

$$\Theta = \{\theta \,|\, \theta \in \mathbb{R}^{\pi_1 JK + (\pi_2+1)J} \times \Gamma\}.$$

In our model, the probability of failing from a risk j is given by a logistic model (Cox, 1970, ch, 7.5). Two cases are considered:

The interior case: $0 < p_i < 1$, $1 \le i \le n$.

In the interior case, the true value is restricted in the interior of the parameter space Θ. Our work will only include one of the most commonly used cases that $\omega(\varrho_i)$ is degenerated to 1 and $\zeta(\xi_j + \rho_{ij}) = \exp{(\xi_j + \rho_{ij})}$, i.e.

$$p_{ij} = \frac{\exp(\xi_j + \beta_j^T y_i)}{1 + \sum_{l=1}^{J} \exp(\xi_l + \beta_l^T y_i)}. \tag{3.2.15}$$

The parameter space is $\Theta_1 = \mathbb{R}^{\pi_1 JK + (\pi_2+1)J}$ and the true value should be

$$\theta_0 = (\alpha_{10}^T, \ldots, \alpha_{J0}^T, \beta_{10}^T, \ldots, \beta_{J0}^T, \xi_{10}^T, \ldots, \xi_{J0}^T).$$

The boundary case: $p_i = 1$, $1 \le i \le n$.

The boundary of the space is denoted by

$$\Theta_2 = \{\theta \,|\, \theta \in \mathbb{R}^{\pi_1 JK + (\pi_2+1)J} \times \Gamma_2\},$$

with

$$\Gamma_2 = \{\gamma \,|\, \gamma^T \in \mathbb{R}^{\pi_3}, \, \omega(\gamma^T w_i) = 0, \text{ for all } 1 \le i \le n\},$$

so that $\sum_{j=1}^{J} p_{ij} = 1$. In this boundary case, let $w_i = 1$ and $\omega(\varrho_i) = \varrho_i$, $1 \le i \le n$. Further assume that ξ_{J0} and β_{J0} are degenerated to 0. Hence p_{ij} can be formulated by

$$p_{ij} = \frac{\exp(\xi_j + \beta_j^T y_i)}{\gamma + 1 + \sum_{l=1}^{J-1} \exp(\xi_l + \beta_l^T y_i)}, \qquad 1 \le j \le J-1 \tag{3.2.16}$$

and

$$p_{iJ} = \frac{1}{\gamma + 1 + \sum_{l=1}^{J-1} \exp(\xi_l + \beta_l^T y_i)}. \tag{3.2.17}$$

The true value of the parameter vector is

$$\theta_0 = (\alpha_{10}^T, \ldots, \alpha_{J0}^T, \beta_{10}^T, \ldots, \beta_{(J-1)0}^T, \xi_{10}^T, \ldots, \xi_{(J-1)0}^T, \gamma_0),$$

which lies on

$$\Theta_3 = \{\theta \,|\, \theta \in \mathbb{R}^{\pi_1 JK + (\pi_2+1)(J-1)} \times \{0\}\}. \tag{3.2.18}$$

And the parameter space under consideration should be

$$\Theta_4 = \{\theta \,|\, \theta \in \mathbb{R}^{\pi_1 JK + (\pi_2+1)(J-1)} \times [0, \infty)\}. \tag{3.2.19}$$

Assume that the following three regularity conditions are satisfied by $f(t; \phi_{ij})$ throughout:

(R1) The support of $f(t; \phi)$ is not influenced by ϕ. It is assumed to be $[0, \infty)$ in our analysis.

(R2) $f(t; \phi)$ is twice-differentiable.

(R3) $\partial f(t; \phi)/\partial\phi$ and $\partial^2 f(t; \phi)/\partial\phi\partial\phi^T$ are continuous and integrable for $t > 0$.

These regularity conditions are commonly required in survival analysis and it is easy to check that most of the distributions, such as the exponential, Weibull and geometry distributions, used in survival analysis satisfy these conditions.

Let $l_n(\theta) = \sum_{i=1}^n l_{ni}(\theta)$ with

$$l_{ni}(\theta) = \sum_{j=1}^{J} \delta_{ij} \log\big(p_{ij} f_{ij}(t_i)\big) + (1 - \delta_i) \log\big(1 - F_i(t_i)\big). \tag{3.2.20}$$

The first derivative of $l_n(\theta)$ with respect to θ is

$$\mathcal{S}_n(\theta) = \frac{\partial l_n(\theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial l_{ni}(\theta)}{\partial \theta},$$

which is a $(\pi_1 JK + (\pi_2+1)J) \times 1$-vector for the interior case and $(\pi_1 JK + (\pi_2+1)(J-1)+1) \times 1$-vector for the boundary case. The negative second derivative matrix of l_n is denoted by

$$\mathcal{F}_n(\theta) = -\frac{\partial^2 l_n(\theta)}{\partial \theta \partial \theta^T} = -\sum_{i=1}^{n} \frac{\partial^2 l_{ni}(\theta)}{\partial \theta \partial \theta^T}, \tag{3.2.21}$$

which is a $(\pi_1 JK + (\pi_2+1)J) \times (\pi_1 JK + (\pi_2+1)J)$ matrix for the interior case and a $(\pi_1 JK + (\pi_2+1)(J-1)+1) \times (\pi_1 JK + (\pi_2+1)(J-1)+1)$ matrix for the boundary case. Then the regularity conditions for $f(t)$ and the link functions imply that

$$\frac{\partial p_{ij}}{\partial \xi_l} = \frac{\partial p_{ij}}{\partial \xi_l} 1, \quad 1 \le i \le n,\, 1 \le j \le J, \tag{3.2.22}$$

$$\frac{\partial p_{ij}}{\partial \beta_l} = \frac{\partial p_{ij}}{\partial \rho_{il}} y_i, \quad 1 \le i \le n,\, 1 \le j \le J, \tag{3.2.23}$$

$$\frac{\partial p_{ij}}{\partial \gamma} = \frac{\partial p_{ij}}{\partial \varrho_i} 1, \quad 1 \le i \le n,\, 1 \le j \le J, \tag{3.2.24}$$

where $1 \le l \le J$ for the interior case and $1 \le l \le J-1$ for the boundary case. And

$$\frac{\partial f_{ij}(t)}{\partial \alpha_{jk}} = \frac{\partial f_{ij}}{\partial \kappa_{ijk}} x_i, \quad 1 \le i \le n,\, 1 \le j \le J,\, 1 \le k \le K, \tag{3.2.25}$$

$$\frac{\partial f_{ij}(t)}{\partial \alpha_{lk}} = 0, \quad 1 \le i \le n,\, 1 \le j \le J,\, j \ne l. \tag{3.2.26}$$

Thus

$$\mathcal{S}_{ni}(\theta) = \frac{\partial l_{ni}(\theta)}{\partial \theta} = H_i \mathscr{S}_i(\theta) \tag{3.2.27}$$

and

$$\mathcal{F}_{ni}(\theta) = -\frac{\partial^2 l_{ni}(\theta)}{\partial \theta \partial \theta^T} = H_i \mathscr{F}_i(\theta) H_i^T \tag{3.2.28}$$

with

$$H_i = \begin{bmatrix} I_{JK} \otimes x_i & 0 & 0 \\ 0 & I_b \otimes y_i & 0 \\ 0 & 0 & I_J \end{bmatrix} \tag{3.2.29}$$

for $1 \leq i \leq n$, where $b = J$ for the interior case and $b = J - 1$ for the boundary case. H_i is a $(\pi_1 JK + \pi_2 b + J) \times (JK + b + J)$ non-stochastic matrix. It should also be noted that $\mathscr{S}_i$ and $\mathscr{F}_i$ are $(JK + b + J) \times 1$ random vector and $(JK + b + J) \times (JK + b + J)$ symmetric random matrix respectively that vary over the interior case and boundary case. As a result, $\mathcal{Q}_n = E[\mathcal{F}_n(\theta_0)]$, the expectation of $\mathcal{F}_n$ with respect to the true value of the parameters, can be expressed by

$$\mathcal{Q}_n = \sum_{i=1}^{n} H_i \mathscr{Q}_i H_i^T, \tag{3.2.30}$$

where $\mathscr{Q}_i = E[\mathscr{F}_i(\theta_0)]$, $1 \leq i \leq n$, are $(JK + b + J) \times (JK + b + J)$ symmetric fixed matrices that also vary over the interior case and boundary case. The elements of $\mathscr{S}_i(\theta)$, $\mathscr{F}_i(\theta)$ and $\mathscr{Q}_i$ are denoted by $s_i^r(\theta)$, $f_i^{rm}(\theta)$ and q_i^{rm} respectively for $1 \leq i \leq n$, $1 \leq r \leq JK + b + J$ and $1 \leq m \leq JK + b + J$, i.e

$$\mathscr{S}_i(\theta) = [s_i^r(\theta)]_{(JK+b+J)\times 1}, \tag{3.2.31}$$

$$\mathscr{F}_i(\theta) = [f_i^{rm}(\theta)]_{(JK+b+J)\times(JK+b+J)} \tag{3.2.32}$$

and

$$\mathscr{Q}_i = [q_i^{rm}]_{(JK+b+J)\times(JK+b+J)}. \tag{3.2.33}$$

Our aim is to estimate the true value of the parameter vector θ_0 by maximizing $l_n(\theta)$ over different subsets of the parameter space in different cases and to derive different test statistics of practical interest. The large sample properties of the proposed estimators and test statistics also need to be verified. In the next section, we will show the sufficient conditions under which a maximum likelihood estimator (MLE) exists local to the true value, and provide its asymptotic distribution. We shall also discuss the sufficient conditions and asymptotic properties of the test statistics.

3.3 Main Results and Examples

Throughout this chapter, I_d denotes an identity matrix of d dimensions, and $N(0, I_s)$ represents a standard normal random vector of s dimensions. For a symmetric matrix C, let $tr(C)$ denotes its trace and $\| C \|_1$ represents the sum of the absolute values of the elements of C. For a positive definite matrix D, $D^{1/2}$ represents the left Cholesky square root (the lower triangular matrix) of D such that $D^{1/2}(D^{1/2})^T = D$.

For a fixed positive constant A, the neighborhood of θ_0 is defined by

$$\mathcal{N}_n(A) = \{\theta : (\theta - \theta_0)^T \mathcal{Q}_n (\theta - \theta_0) \leq A^2\}, \tag{3.3.34}$$

where $\mathcal{Q}_n$ is defined by (3.2.30). To make inferences on our general mixture model, the following assumptions about the covariates and the expectations of the score statistic and information matrix are needed:

(B1) $\lim_{n\to\infty} \sum_{i=1}^{n} \left(tr\{H_i^T \mathcal{Q}_n^{-1} H_i\}\right)^{3/2} = 0.$

(B2) For all $1 \leq i \leq n$, $1 \leq r \leq JK + b + J$ and $1 \leq m \leq JK + b + J$, we have

$$Var\left(\{f_i^{rm}(\theta_0)\}\right) \leq M, \tag{3.3.35}$$

where f_i^{rm} are defined by (3.2.32) and M is a positive constant.

(B3) For all $1 \le i \le n$, $1 \le r \le JK + b + J$ and $1 \le m \le JK + b + J$,

$$E\{\sup_{\theta \in \mathcal{N}_n(A)} |f_i^{rm}(\theta) - f_i^{rm}(\theta_0)|\} \le M' A\big(tr\{H_i^T \mathcal{Q}_n^{-1} H_i\}\big)^{1/2} \tag{3.3.36}$$

for all $A > 0$. Like $(B2)$, f_i^{rm} here are defined by (3.2.32) and M' is also a positive constant.

(B4) There exists $M'' > 0$ such that

$$E\big[\big(s_i^r(\theta_0)\big)^4\big] \le M''$$

for all $1 \le i \le n$ and $1 \le r \le JK + b + J$, where $s_i^r(\theta)$ is given by (3.2.31).

Assumption (B1) implies that $\mathcal{Q}_n$ is non-singular and also ensures that the covariates do not degenerate to a lower dimensional subspace for n large enough. Assumption (B2) - (B4) depend on specific models and can be checked by the usual calculus methods. The assumptions are satisfied by most of the commonly used survival distributions, such as exponential, Weibull and Gamma, under some mild assumptions.

The next theorem shows the existence of an MLE of the parameter vector local to an interior true value, and provides its asymptotic distribution.

Theorem 3.3.1 *Suppose that assumptions (B1)-(B4) hold. Then, with probability approaching 1 (WPA1), as $n \to \infty$ there exists an interior maximizer $\hat{\theta}_n^{(1)}$ in Θ_1 such that:*

(i) $\hat{\theta}_n^{(1)}$ is unique in a neighborhood of θ_0;

(ii) $\lim_{n\to\infty} \hat{\theta}_n^{(1)} = \theta_0$;

(iii) as $n \to \infty$, we have

$$(Q_n^{1/2})^T(\hat{\theta}_n^{(1)} - \theta_0) \xrightarrow{D} N(0, I_{\pi_1 JK+\pi_2 J+J}) \tag{3.3.37}$$

and

$$\left(\mathcal{F}_n^{1/2}(\hat{\theta}_n^{(1)})\right)^T(\hat{\theta}_n^{(1)} - \theta_0) \xrightarrow{D} N(0, I_{\pi_1 JK+\pi_2 J+J}), \tag{3.3.38}$$

where $\mathcal{Q}_n$ is defined by (3.2.30) and $\mathcal{F}_n$ is defined by (3.2.21).

Theorem 3.3.1 shows the existence and large sample properties of the maximum likelihood estimator $\hat{\theta}_n^{(1)}$ of the interior true point θ_0 in the parameter space. In the theorem, (ii) indicates that $\hat{\theta}_n^{(1)}$ is a consistent estimator and (iii) means that if $\hat{\theta}_n^{(1)}$ is normed by the left Cholesky square root of sample information matrix, it has an asymptotic standard normal distribution. Thus the covariance matrix of $\hat{\theta}_n^{(1)}$ can be approximated by the inverse of the estimated information matrix, which can be obtained directly by $F_n(\hat{\theta}_n^{(1)})$.

The next part is about the interior hypotheses testing.

Let S_r be an r-dimensional subspace of $\mathbb{R}^{(\pi_1 JK+\pi_2 J+J)}$ with $0 \leq r < \pi_1 JK+\pi_2 J+J$, and let θ^* be any specified point in the interior of Θ_1. The null hypothesis of our interest is

$$H_0^{(1)}\colon \theta_0 \in S_r + \theta^*. \tag{3.3.39}$$

The alternative hypothesis with respect to $H_0^{(1)}$ is

$$H_1^{(1)}\colon \theta_0 \in \Theta_1. \tag{3.3.40}$$

Thus the restricted parameter space Ω_r under $H_0^{(1)}$ is

$$\Omega_r = (S_r + \theta^*) \cap \Theta_1 = (S_r + \theta^*) \cap \mathbb{R}^{(\pi_1 JK+\pi_2 J+J)}, \tag{3.3.41}$$

while the parameter space under unrestricted interior alternative is still

$$\Theta_1 = \mathbb{R}^{(\pi_1 JK + \pi_2 J + J)}.$$

We have the following theorems.

Theorem 3.3.2 *Suppose that assumptions (B1)-(B4) hold. Then with probability approaching 1, there exists a maximizer $\hat{\theta}_n^{(2)}$ over Ω_r, which is unique and consistent in a neighborhood of θ_0 in Ω_r.*

Theorem 3.3.2 guarantees that there is a maximum likelihood estimator $\hat{\theta}_n^{(2)}$ of θ_0 in the restricted parameter space Ω_r under $H_0^{(1)}$. The next theorem is about the asymptotic distribution of the test statistic for $H_0^{(1)}$.

Theorem 3.3.3 *Suppose that assumptions (B1)-(B4) hold. Then the deviance statistic follows an asymptotic Chi-square distribution with $\pi_1 JK+\pi_2 J+J-r$ degrees of freedom, i.e.*

$$d_n^{(1)} = -2\big(l_n(\hat{\theta}_n^{(2)}) - l_n(\hat{\theta}_n^{(1)})\big) \xrightarrow{D} \chi^2_{\pi_1 JK+\pi_2 J+J-r}, \tag{3.3.42}$$

where $\theta_n^{(1)}$ is defined in Theorem 3.3.1 and $\theta_n^{(2)}$ is defined in Theorem 3.3.2.

Theorem 3.3.3 provides a likelihood ratio test not only for a hypothesis where there are some specified parameters but also for hypotheses about relationships among the parameters. For instance, problems 1, 3 and 4 proposed in Chapter 1 could be answered via different null hypotheses according to specific problems. We list the tests below:

1. Recall that problem 1 in Chapter 1 is about the value of p_{ij}. Consider the hypothesis

$$H_0^{(11)}\colon \xi_1 = \xi_1^*, \ldots, \xi_J = \xi_J^*,\ \beta_1 = \beta_1^*, \ldots, \beta_J = \beta_J^*, \tag{3.3.43}$$

where $\theta^* = [\xi_1^{*T}, \ldots, \xi_J^{*T},\ \beta_1^{*T}, \ldots, \beta_J^{*T}]^T$ is a $(\pi_2 J + J)$-dimensional constant vector and $S_r = \mathbb{R}^{(\pi_1 JK)}$. Thus the deviance statistic $d_n^{(11)}$ asymptotically follows a Chi-square distribution with $\pi_2 J + J$ degrees of freedom. This test can be applied to test whether the regression coefficients of the covariates equal to specific values. Especially, by letting $\theta^* = [0]_{(\pi_2 J + J)\times 1}$, we are able to test whether the covariates have influence on the failure rates with respect to the risks.

2. Problem 3 is related to the significance levels of the regression coefficients for the covariates x_i and y_i. For instance, to test whether $x_{i(1)}, \ldots, x_{i(r_1)}$ and $y_{i(1)}, \ldots, y_{i(r_2)}$ can be removed from the covariates under consideration, we should test

$$H_0^{(12)}\colon \alpha_{jk(1)} = \cdots = \alpha_{jk(r_1)} = 0, \quad \beta_{j(1)} = \cdots = \beta_{j(r_2)} = 0, \tag{3.3.44}$$

where $[x_{i(1)}, \ldots, x_{i(r_1)}]$ is a subvector of x_i and $[y_{i(1)}, \ldots, y_{i(r_2)}]$ is a subvector of y_i with $1 \le r_1 \le \pi_1$ and $1 \le r_2 \le \pi_2$. Thus $H_0^{(12)}$ implies that $S_r = \mathbb{R}^{((\pi_1 - r_1)JK + (\pi_2 - r_2 + 1)J)}$ and $\theta^* = 0$. According to Theorem 3.3.2, the asymptotic distribution of $d_n^{(12)}$ for this test is $\chi^2_{r_1 JK + r_2 J}$.

3. Another type of questions of our interest is whether the hazard functions with respect to different types of risks are proportional to each other. In other words, are f_{ij}'s proportional to each other for fixed i? The mull hypotheses for such tests depends on the formulations of f_{ij}'s. As a special case, the differences among the c.d.f.'s conditional on the risks can be tested. The hypothesis can be formulated

as

$$H_0^{(13)}\colon \alpha_{jk} = 0, \quad k = 1, \ldots, K,\ j = 1, \ldots, J. \tag{3.3.45}$$

In this situation $S_r = \mathbb{R}^{(\pi_2+1)J}$ and $\theta^* = 0$. By Theorem 3.3.2, the asymptotic distribution of $d_n^{(13)}$ is $\chi^2_{\pi_1 JK}$.

Theorems 3.3.1, 3.3.2 and 3.3.3 generalize the result of Maller and Zhou (2002), which are about similar problems in the absence of covariates. The interior setting indicates that there are individuals that are not susceptible to the risks $\{1, \ldots, J\}$. In other words, the failure causes under consideration are not exhaustive. This is reasonable if the existence of immunes is previously known. In practice, however, it is difficult to identify the existence of immunes from real life data because of censoring. Actually, this issue has been of increasing interest recently. To investigate this problem, we may wish to develop a test of the hypothesis that the J risks of interest are exhaustive, i.e. all failure causes to which the individuals are susceptible are included in our investigation.

In the boundary setting, consider the hypothesis

$$H_0^{(2)}\colon \gamma_0 = 0. \tag{3.3.46}$$

Thus the restricted parameter space under $H_0^{(2)}$ is Θ_3 and the unrestricted parameter space is Θ_4. The following theorems are about the point estimation under the null hypothesis of boundary.

Theorem 3.3.4 *Suppose that assumptions (B1)-(B4) hold. Then*

(i) With probability approaching 1, there exists a maximizer $\hat{\theta}_n^{(3)}$ over Θ_3, which is unique and consistent in a neighborhood of θ_0 in Θ_3.

(ii) With probability approaching 1, there exists a maximizer $\hat{\theta}_n^{(4)}$ over Θ_4, which is unique and consistent in a neighborhood of θ_0 in Θ_4.

Theorem 3.3.4 shows the existence, consistency, and asymptotic normality of the maximum likelihood estimators $\hat{\theta}_n^{(3)}$ over Θ_3 and $\hat{\theta}_n^{(4)}$ over Θ_4 for true point θ_0. From the work of Vu and Zhou (1997), the next theorem can be developed. It plays an important role in our investigation because it provides a method to test the exhaustive of the failure causes under consideration.

Theorem 3.3.5 *Suppose that assumptions (B1)-(B4) hold. Then the deviance statistic is asymptotically distributed as a mixture of a Chi-square distribution with 1 degree of freedom and 0, i.e.*

$$d_n^{(2)} = -2\big(l_n(\hat{\theta}_n^{(3)}) - l_n(\hat{\theta}_n^{(4)})\big) \xrightarrow{D} N^2 I\{N \leq 0\}, \tag{3.3.47}$$

where N is the standard normal distribution and $\theta_n^{(3)}$ and $\theta_n^{(4)}$ are defined in Theorem 3.3.4.

Theorem 3.3.5 is an extension of the work of Maller and Zhou (2002) in which a boundary hypothesis test is discussed in the absence of covariate information. This theorem allows us to verify whether there are individuals who are immune to all of the causes under consideration by comparing the deviance value to the critical value. We can accommodate data with suitable models according to the testing results, i.e. the interior models are reasonable if $H_0^{(2)}$ is rejected, otherwise the boundary model can be used to fit the data. The limiting distribution of the deviance $d_n^{(2)}$ is a 50-50 mixture distribution of a Chi-square random variable with 1 degree of freedom and a point mass at 0. As $P\{N^2 I\{N \leq 0\} > c\} = 0.5P\{\chi^2 > c\}$ for any non-negative constant c, the critical value of $N^2 I\{N \leq 0\}$ can be easily calculated from χ^2, i.e. its 95th percentile is 2.71 and the 99.5th percentile is 6.635. For the details of the application of the 50-50 mixture Chi-square distribution and 0, see Zhou and Maller (1995) and Maller and Zhou (1996).

An important issue to be emphasized is that if the boundary model is chosen by the result of $H_0^{(2)}$, we can still carry out the test $H_0^{(1)}$. In boundary model, under $H_0^{(1)}$ the deviance statistic asymptotically follows a Chi-square distribution with $\pi_1 JK + (\pi_2 + 1)(J-1) - r$ degrees of freedom.

3.4 Simulations

A simulation is carried out to illustrate the performance of our approach. A commonly used exponential mixture model is investigated in our study, where the survival time of individual i given that $D_i = j$ is assumed to be exponentially distributed and $P\{D_i = j\} = p_{ij}$, i.e. for $t > 0$,

$$f_{ij}(t) = \lambda_{ij} \exp(-\lambda_{ij} t), \qquad 1 \leq i \leq n,\ 1 \leq j \leq J \tag{3.4.48}$$

and

$$F_i(t) = \sum_{j=1}^{J} p_{ij}\big(1 - \exp(-\lambda_{ij} t)\big), \qquad 1 \leq i \leq n,$$

where λ_{ij}'s are linked to x_i, the sub-vector of the covariates of individual i, by

$$\lambda_{ij} = \exp(\alpha_j^T x_i), \qquad 1 \leq i \leq n,\ 1 \leq j \leq J. \tag{3.4.49}$$

The failure rate with respect to a specific risk j of individual i is related to y_i, another sub-vector of the covariate vector of individual i, by (3.2.15) for the interior case and by (3.2.16) and (3.2.17) with $\gamma = 0$ for boundary case. The covariates x_i and y_i are set to be one dimensional in the simulation. As a result, the parameters to be estimated for the interior case are

$$\theta_1 = (\alpha_1,\, \alpha_2,\, \xi_1,\, \xi_2,\, \beta_1,\, \beta_2)^T$$

and the parameters to be estimated for the boundary case are

$$\theta_2 = (\alpha_1,\, \alpha_2,\, \xi,\, \beta)^T$$

with true values

$$\theta_{10} = (-1.0000,\, -2.0000,\, 0.9016,\, 0.8188,\, -0.3561,\, 0.1099)^T$$

and

$$\theta_{20} = (-1.0000,\, -2.0000,\, 0.9016,\, -0.3561)^T$$

Both x_i and y_i are random variables with discrete uniform distributions. The value of x_i is set to be one of 0.4321 and 0.8588 and the value of y_i is assumed to be one of 0.1398 and 0.3968. Thus the individuals are classified into four groups by the values of (x_i, y_i). The censoring time c_i for individual i is uniformly distributed in $[20, 30]$.

Two sets of data for the interior case and boundary case are generated by Matlab from the mixture exponential distributions specified above respectively. The sample sizes are taken as N=100, 200, 300, 400, 500, 600, 700, 800, 900 and 1000 with 100 repetitions. We fit each dataset with both models of interior case and boundary case and compute the maximizer of the log-likelihood functions as the estimates of the parameters. Take the interior case as an example, we first fit the data generated by the models defined by (3.2.4), (3.2.15), (3.4.48) and (3.4.49) to assess the performances of the inference methods by comparing the estimates and the true value θ_{10}. Then we approximate the data by the models specified by (3.2.4), (3.2.16), (3.2.17), (3.4.48) and (3.4.49) to assess the performances of the two models for the interior data by comparing the fitted curves of the overall distribution function generated by both models with the true c.d.f. curve. For the data of boundary case, simulation is conducted in a similar way. Results of the point estimates and fitted c.d.f.'s are listed in Table 3.1 - 3.4 and Figure 3.1 - 3.8 respectively.

In this simulation, the estimators are defined to be the mean of the maximum likelihood estimates developed by all the replicates. Table 3.1 and Table 3.2 list the results of the boundary data and Table 3.3 and Table 3.4 are for interior data, where STD denotes the standard deviation and N represents the sample size. In addition, we plot in Fig.3.1 - Fig.3.4 the overall c.d.f. curves of the boundary data based on true parameters, estimates given by boundary model and estimates given by interior model. Similarly, Fig.3.5 - Fig.3.8 are for the comparison of of the overall c.d.f. curves of the interior data based on true parameters, estimates developed by interior model and estimates developed by boundary model. The estimates used in the figures are generated by one replicate with the sample size N=100.

From Table 3.1 and Table 3.3, we can see that the estimates are reasonably accurate and the accuracy improves as N grows. A larger sample size may be required to achieve reasonable accuracy when the number of parameters increases. It is not surprising to see from Fig.3.1 - Fig.3.4 that the fitness of interior model applying to boundary data is not bad as the interior model allows $1-\sum_{j=1}^{J} p_{ij}$ to be almost 0. But we cannot replace the boundary model by interior model in applications because the standard deviations of the estimates are extremely large when boundary data are fitted by interior model. In contrary, Fig.3.5 - Fig.3.8 show that the fitness of boundary model applying to interior data is poor as the boundary model ignores the existence of immunes. That explains why it is important to identify the existence of immunes in survival analysis.

Table 3.1: Results of Boundary Data Fitted by Boundary Model

	Parameters	α_1	α_2	ξ	β
N	True values	-1.0000	-2.0000	0.9136	-0.3561
100	Estimate	-0.9856	-1.9876	1.0303	-0.3733
	STD	0.1911	0.2098	1.5352	0.4809
200	Estimate	-0.9945	-2.0005	0.9284	-0.3484
	STD	0.1279	0.1589	1.1136	0.3328
300	Estimate	-0.9847	-2.0000	0.9056	-0.3459
	STD	0.1141	0.1278	0.9166	0.2924
400	Estimate	-0.9763	-1.9997	0.8999	-0.3486
	STD	0.1015	0.1044	0.8250	0.2515
500	Estimate	-0.9800	-2.0094	0.8991	-0.3509
	STD	0.1024	0.0990	0.6782	0.2000
600	Estimate	-0.9805	-2.0079	0.8904	-0.3480
	STD	0.0851	0.0897	0.6812	0.2002
700	Estimate	-0.9841	-2.0056	0.8900	-0.3494
	STD	0.0803	0.0843	0.5256	0.1584
800	Estimate	-0.9900	-2.0046	0.9257	-0.3638
	STD	0.0828	0.0739	0.5251	0.1545
900	Estimate	-0.9934	-2.0060	0.9180	-0.3606
	STD	0.0664	0.0742	0.4697	0.1463
1000	Estimate	-0.9976	-2.0110	0.9183	-0.3594
	STD	0.0696	0.0804	0.4332	0.1395

Table 3.2: Results of Boundary Data Fitted by Interior Model

N	Parameters	α_1	α_2	ξ_1	ξ_2	β_1	β_2
100	Estimate	-0.9853	-1.9644	125.9748	127.0195	15.1838	14.8031
	STD	0.1908	0.2089	399.0192	399.0058	64.0790	64.0868
200	Estimate	-0.9942	-1.9864	76.2962	77.2323	27.3727	27.0206
	STD	0.1277	0.1584	340.1610	340.2234	94.7250	94.7439
300	Estimate	-0.9847	-1.9877	58.3858	59.2943	34.6003	34.2522
	STD	0.1140	0.1312	490.2110	490.0754	156.5305	156.4826
400	Estimate	-0.9763	-1.9922	13.5408	14.4490	47.1526	46.8007
	STD	0.1015	0.1079	445.9458	445.8482	159.4396	159.4045
500	Estimate	-0.9798	-2.0034	59.4537	60.3564	31.4325	31.0799
	STD	0.1024	0.1016	368.0414	368.0962	93.8851	93.8886
600	Estimate	-0.9804	-2.0028	-42.4383	-41.5458	62.1624	61.8132
	STD	0.0850	0.0896	471.6789	471.6527	175.1202	175.1130
700	Estimate	-0.9841	-2.0019	21.9784	22.8692	37.6823	37.3323
	STD	0.0803	0.0850	224.9419	224.9444	72.5659	72.5632
800	Estimate	-0.9900	-2.0023	41.5852	42.5125	43.9957	43.6312
	STD	0.0828	0.0753	455.3250	455.3576	120.8669	120.8682
900	Estimate	-0.9934	-2.0032	-19.1711	-18.2484	62.9437	62.5812
	STD	0.0664	0.0739	603.2434	603.2077	203.2201	203.1933
1000	Estimate	-0.9976	-2.0088	-77.9795	-77.0613	102.4376	102.0778
	STD	0.0696	0.0808	908.6678	908.6989	287.0666	287.0690

Table 3.3: Results of Interior Data Fitted by Interior Model

	Parameters	α_1	α_2	ξ_1	ξ_2	β_1	β_2
N	True values	-1.0000	-2.0000	0.9016	0.8188	-0.3561	0.1099
100	Estimate	-0.9382	-1.9837	0.9234	0.7867	-0.3499	0.1191
	STD	0.2913	0.2505	2.2162	2.0976	0.6383	0.5843
200	Estimate	-0.9682	-1.9807	0.7828	0.7483	-0.3071	0.1397
	STD	0.1869	0.1797	1.5322	1.3359	0.4536	0.3718
300	Estimate	-0.9858	-1.9865	0.9340	0.8805	-0.3611	0.0980
	STD	0.1476	0.1460	1.2608	0.9118	0.3697	0.2349
400	Estimate	-0.9929	-1.9834	0.9326	0.8737	-0.3711	0.0926
	STD	0.1386	0.1177	1.1489	0.8232	0.3361	0.2384
500	Estimate	-0.9971	-1.9830	0.8466	0.8528	-0.3421	0.0952
	STD	0.1196	0.1178	1.0123	0.8095	0.2837	0.2296
600	Estimate	-0.9997	-1.9771	0.8245	0.8310	-0.3457	0.0918
	STD	0.1137	0.1056	0.8707	0.6632	0.2522	0.1965
700	Estimate	-1.0038	-1.9765	0.8522	0.8429	-0.3512	0.0906
	STD	0.0979	0.0970	0.8472	0.6449	0.2419	0.1802
800	Estimate	-0.9948	-1.9816	0.8545	0.7989	-0.3503	0.1074
	STD	0.0985	0.0868	0.8141	0.6275	0.2322	0.1853
900	Estimate	-0.9979	-1.9834	0.8507	0.7782	-0.3505	0.1139
	STD	0.0953	0.0782	0.7146	0.6074	0.2030	0.1702
1000	Estimate	-1.0011	-1.9854	0.8110	0.7556	-0.3384	0.1203
	STD	0.0856	0.0808	0.6586	0.5605	0.1930	0.1694

Table 3.4: Results of Interior Data Fitted by Boundary Model

N	Parameters	α_1	α_2	ξ	β
100	Estimate	-0.9377	-5.0508	-0.4508	1.1124
	STD	0.2906	0.4492	1.7760	0.5401
200	Estimate	-0.9680	-5.0084	-0.3386	1.0737
	STD	0.1866	0.2923	1.2745	0.3955
300	Estimate	-0.9856	-5.0181	-0.4096	1.1003
	STD	0.1474	0.2887	1.1142	0.3380
400	Estimate	-0.9927	-5.0218	-0.4222	1.1089
	STD	0.1383	0.2207	1.0398	0.2982
500	Estimate	-0.9968	-5.0319	-0.3489	1.0811
	STD	0.1195	0.2124	0.8547	0.2442
600	Estimate	-0.9994	-5.0385	-0.3432	1.0830
	STD	0.1136	0.1855	0.7578	0.2194
700	Estimate	-1.0036	-5.0396	-0.3621	1.0870
	STD	0.0977	0.1629	0.7046	0.2053
800	Estimate	-0.9946	-5.0322	-0.3907	1.0959
	STD	0.0984	0.1500	0.6816	0.1953
900	Estimate	-0.9977	-5.0302	-0.3978	1.0992
	STD	0.0951	0.1481	0.6061	0.1760
1000	Estimate	-1.0009	-5.0308	-0.3731	1.0910
	STD	0.0854	0.1313	0.5504	0.1662

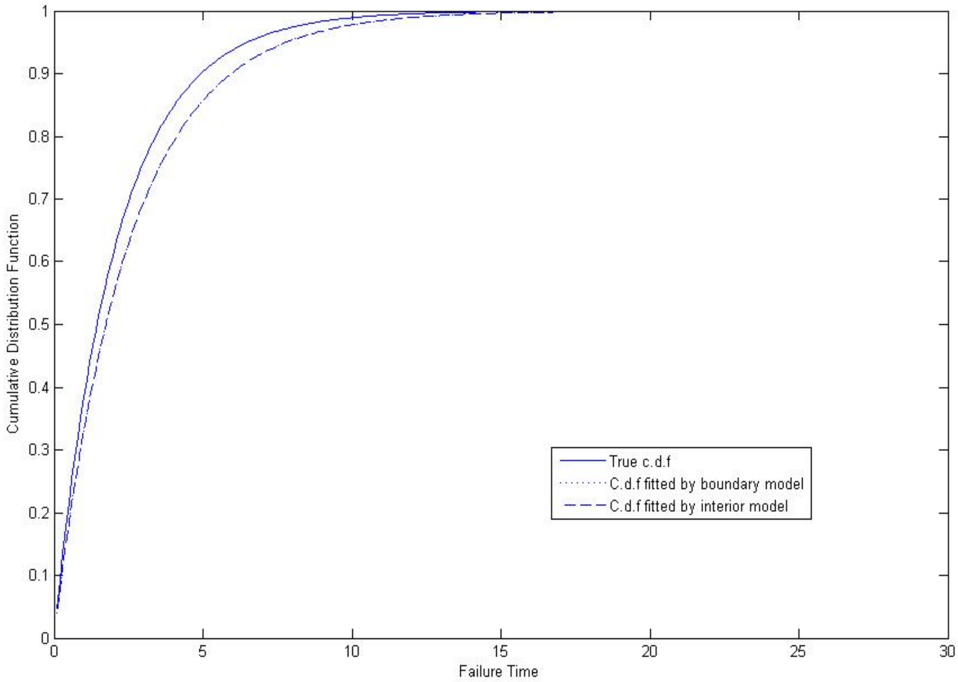

Figure 3.1: Comparison of C.d.f.'s for Boundary Data with x=0.4321 and y=0.1398

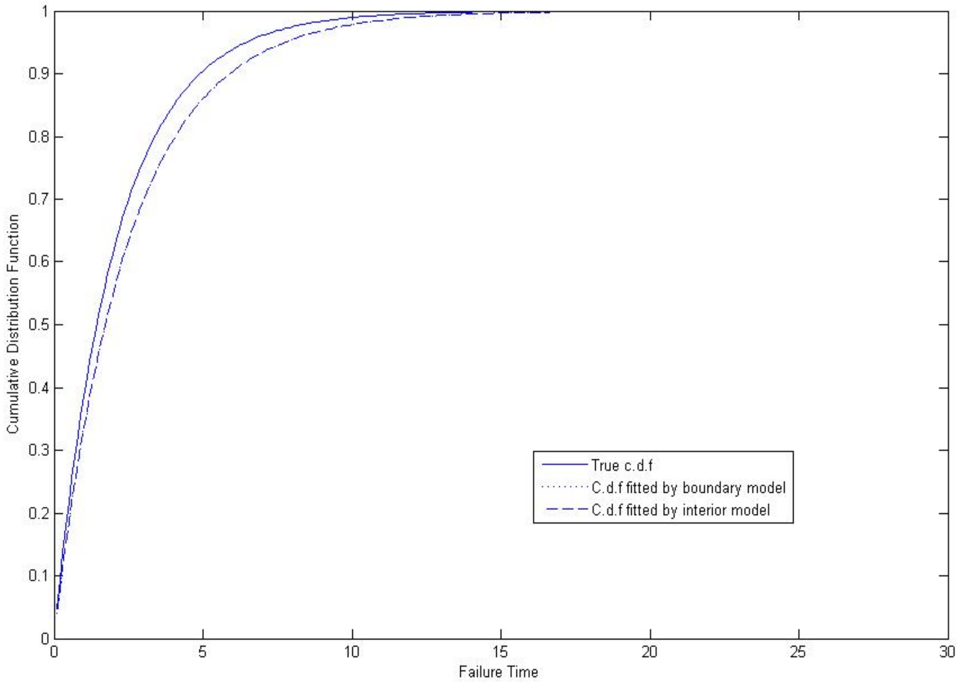

Figure 3.2: Comparison of C.d.f.'s for Boundary Data with x=0.4321 and y=0.3968

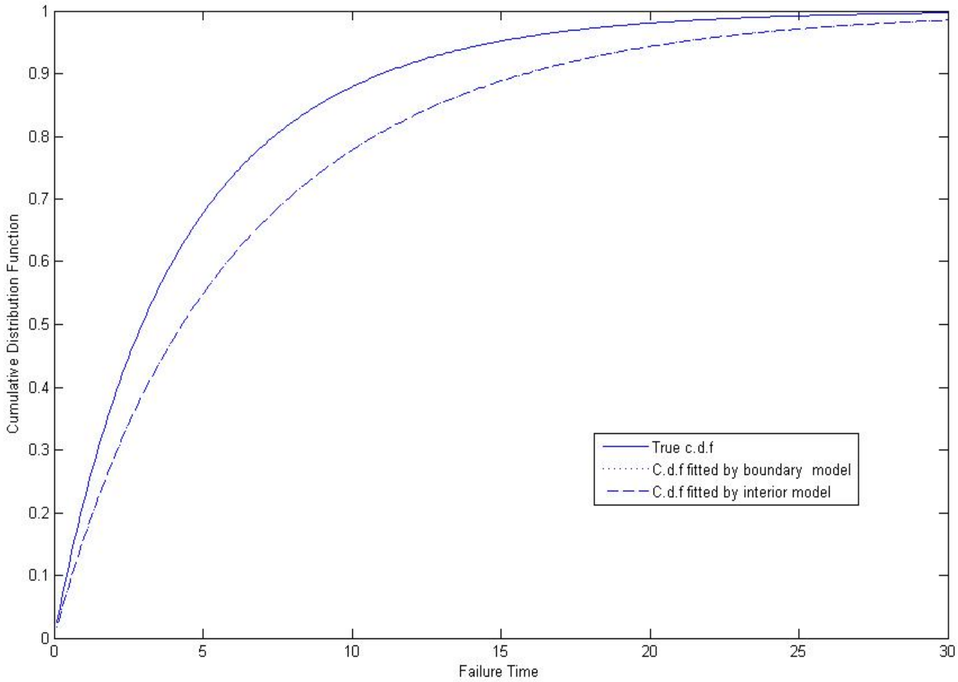

Figure 3.3: Comparison of C.d.f.'s for Boundary Data with x=0.8588 and y=0.1398

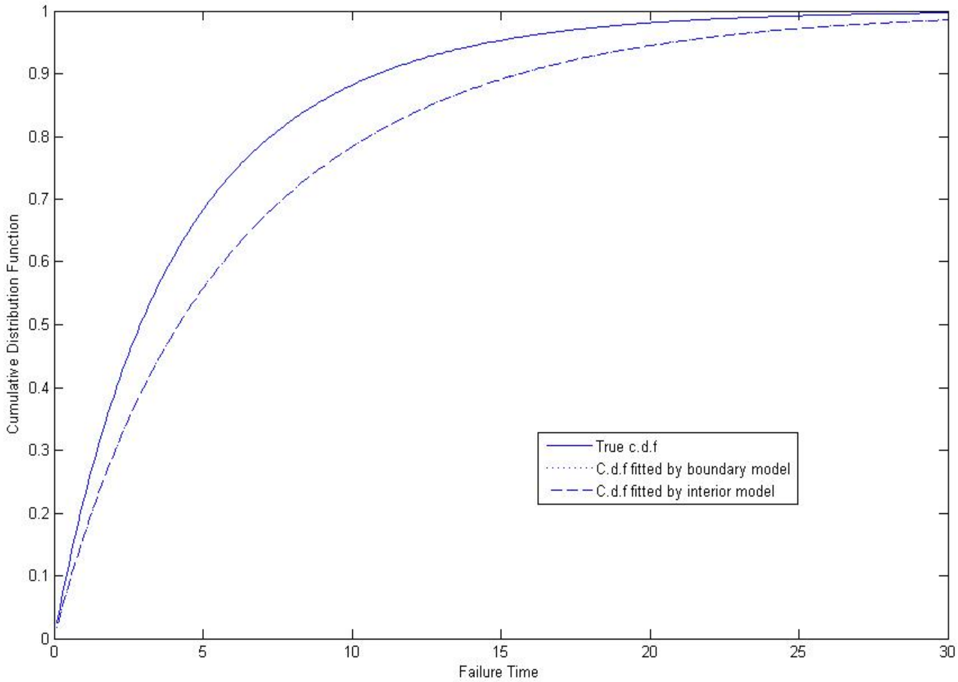

Figure 3.4: Comparison of C.d.f.'s for Boundary Data with x=0.8588 and y=0.3968

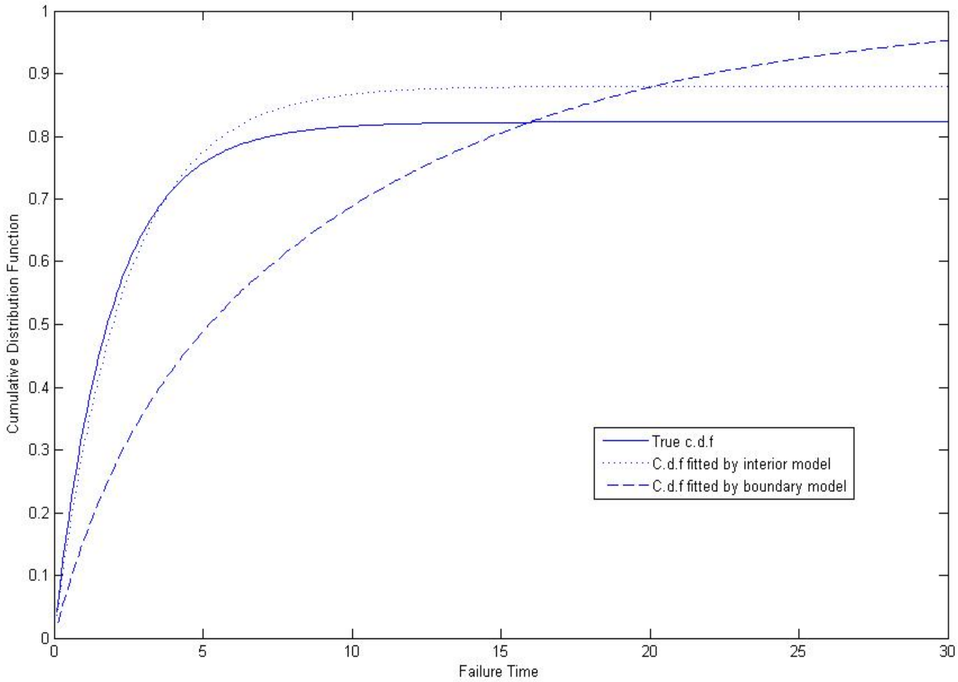

Figure 3.5: Comparison of C.d.f.'s for Interior Data with x=0.4321 and y=0.1398

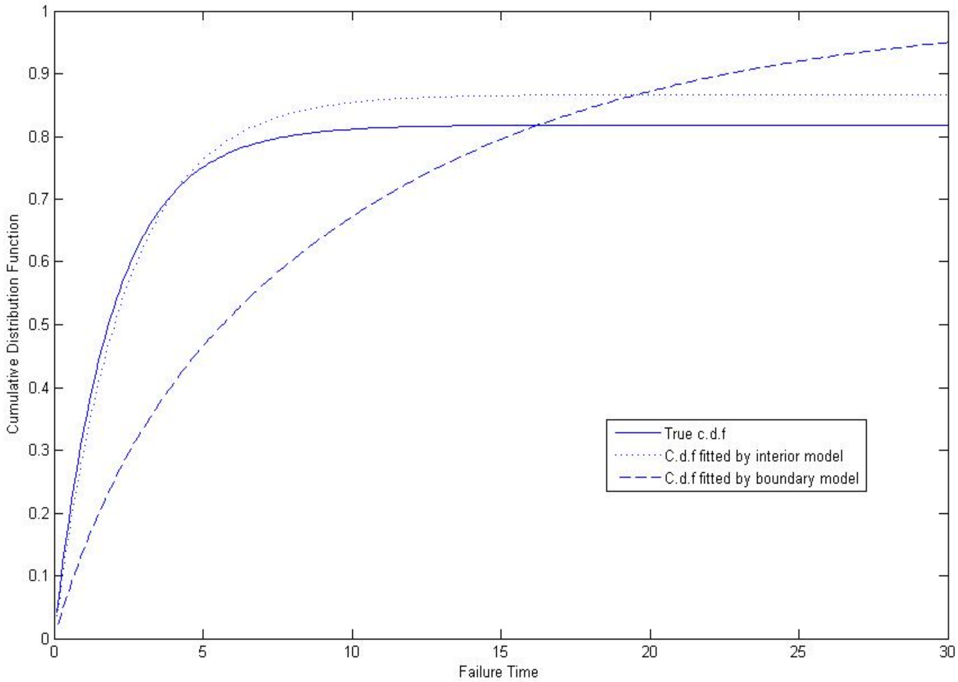

Figure 3.6: Comparison of C.d.f.'s for Interior Data with x=0.4321 and y=0.3968

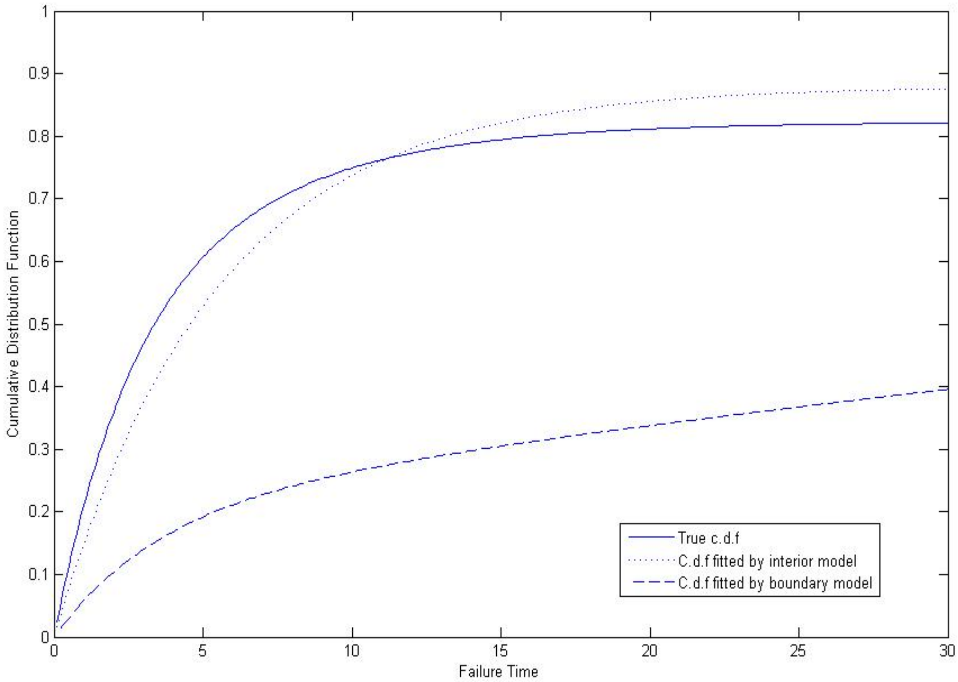

Figure 3.7: Comparison of C.d.f.'s for Interior Data with x=0.8588 and y=0.1398

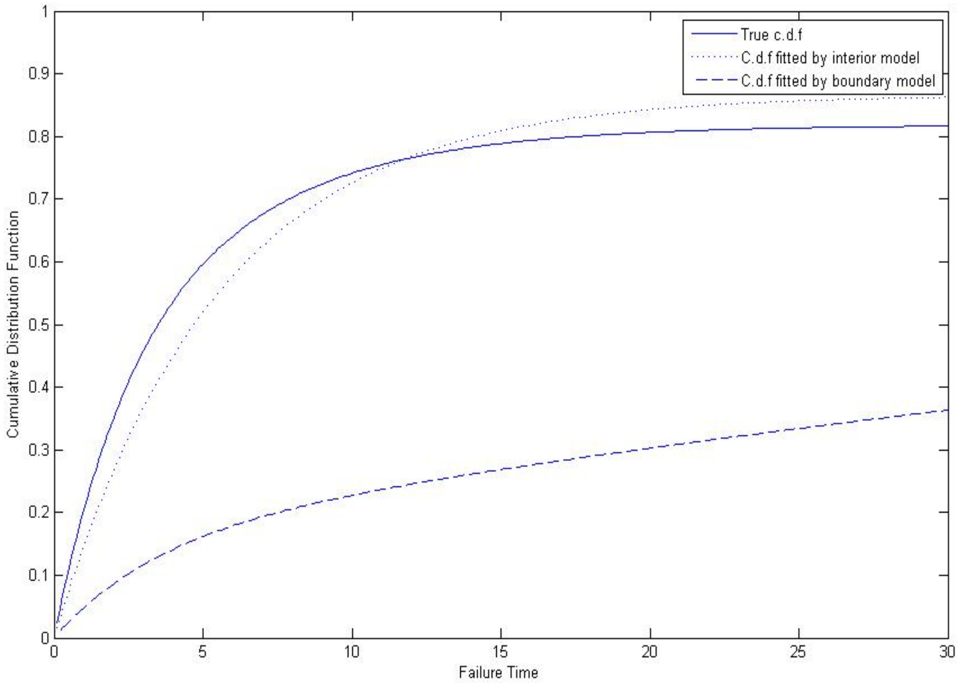

Figure 3.8: Comparison of C.d.f.'s for Interior Data with x=0.8588 and y=0.3968

3.5 Concluding Remarks

In this chapter, a general mixture model is proposed to analyze competing risks data with covariates and possible immunes. Our approach is based on the assumption of independent censoring. The maximum likelihood estimators and their asymptotic properties are obtained and investigated under some fairly general conditions. The main findings are the basic properties of consistency and asymptotic normality hold in both interior and boundary cases and the large-sample 50-50 chi-square distribution for the boundary hypothesis test that none of the population is immune to the causes under consideration. The theoretical results provide the basic large sample foundations for analysis of competing risks data using the mixture model.

The simulation study indicates that the proposed model and estimation procedures produce efficient estimators for exponential mixture models. Further specific distribution functions that are commonly used in survival analysis could be investigated in a similar way. The idea is to check the sufficient conditions required in (B1)-(B4) under different assumptions for different models with some technical details.

The boundary hypothesis test proposed in this chapter is essentially a test for whether the failure causes of interest include all causes of failure since there may be unobserved potential causes. If the existence of extra causes are tested, an improper mixture model should be accommodated to approximate the data. Otherwise, the model should be restricted on the boundary of the parameter space.

The identifiability problem of immunes is related to the duration of follow-up. Theoretically, the ideal follow-up should be ∞, but this is impossible in practice. This issue has been investigated by Maller and Zhou (1992; 1996, p.37). The conditions and a suggested test for sufficiency of follow-up are developed in their work. We expect that Maller and Zhou's methods could be generalized to investigate the problems of sufficient follow-up in competing risks data with possible immunes.

Future work also includes a study of nonparametric approach in the framework of general mixture model. Many techniques have been discussed in the literature. For the analysis of covariates, see Kalbfleisch and Prentice (1980, p.183), Cox and Oakes (1984, p.143). For a discussion of competing risks data via non-parametric approaches, see Lagakos, Sommer and Zelen (1978) and Andersen et al. (1993). The estimation and testing methods could be developed based on the previous work in future.

3.6 Proofs

Our proofs will mainly take advantage of the results in Vu and Zhou (1997). Here, the deviance statistic is defined to be the difference of the values of an estimating function evaluated at its local maximizer over two different subsets of the parameter space. The large sample properties of the estimators and deviance are investigated under some natural and fairly mild conditions. The regularity conditions of density functions with respect to specific risks and link functions guarantee that the log likelihood functions are second order differentiable. For more details, see conditions $(A1)$, $(A2)$, $(A2')$ and $(A3)$ of Vu and Zhou(1997) and the proof of Maller and Zhou (2002). Conditions $(B1)$-$(B5)$ of Vu and Zhou (1997) concern asymptotic behavior of the first and second derivatives of log-likelihood functions and their expectations. We restate the conditions as follows:

(C1) $E[\mathcal{S}_n(\theta_0)] = 0$ and $E[\mathcal{S}_n(\theta_0)\mathcal{S}_n(\theta_0)^T] = E[\mathcal{F}_n(\theta_0)]$ is finite.

(C2) $\lambda_{min}(\mathcal{Q}_n) \to \infty$ as $n \to \infty$, i.e. $\mathcal{Q}_n$ is positive definite as n large enough, where λ_{min} denotes the smallest eigenvalue of a matrix.

(C3) For any positive constant A, as $n \to \infty$

$$\sup_{\theta \in \mathcal{N}_n(A)} ||\mathcal{Q}_n^{-1/2}\mathcal{F}_n(\theta)(\mathcal{Q}_n^{-1/2})^T - I||_1 \to 0.$$

(C4) Condition $C2$ holds and

$$\mathcal{Q}_n^{-1/2}\mathcal{S}_n(\theta_0) \xrightarrow{D} N(0, I).$$

It should be noted that in condition $(C3)$ we take $\mathcal{Q}_n = G_n$ and $V = I$ in Vu and Zhou's original conditions (B1)-(B4) and thus make $B4$ redundant.

For the proofs we need the formulations of $\mathcal{S}_n(\theta)$ and $\mathcal{F}_n(\theta)$.

Lemma 3.6.1 *Assume that the regularity conditions* $(R1)$-$(R3)$ *are satisfied by* $f(t;\phi_{ij})$ *for* $1 \le i \le n$ *and* $1 \le j \le J$, *then condition* $(C1)$ *holds.*

Proof. The first derivative of l_{ni}, the contribution to the log-likelihood function by individual i is

$$\frac{\partial l_{ni}(\theta)}{\partial \theta} = \sum_{j=1}^{J} \delta_{ij} \frac{\partial\{p_{ij} f_{ij}(t_i)\}}{p_{ij} f_{ij}(t_i)\partial\theta} + (1-\delta_i)\frac{\partial S_i(t_i)}{S_i(t_i)\partial\theta}.$$

Note that for any measurable function $M(\cdot)$ on $\mathbb{R}$, we have

$$\begin{aligned} E[\delta_{ij} M(t_i)] &= E[\delta_i M(t_i))|D_i = j]P(D_i = j) \\ &= p_{ij0} E[E[\delta_i M(t_i)|u_i, D_i = j)]] \\ &= p_{ij0} E\left\{\int_0^{u_i} M(t) dF_{ij0}(t)\right\} \end{aligned} \tag{3.6.50}$$

and

$$\begin{aligned} E[(1-\delta_i)M(t_i)] &= E[E[(1-\delta_i)M(t_i)|u_i]] \\ &= E\left\{\int_{t=u_i}^{\infty} M(u_i) dF_{i0}(t)\right\} \\ &= E\left\{M(u_i)S_{i0}(u_i)\right\}. \end{aligned} \tag{3.6.51}$$

Thus by the regularity conditions we get

$$
\begin{aligned}
E\left\{\frac{\partial l_{ni}(\theta)}{\partial\theta}\right\}_{\theta=\theta_0} =& \sum_{j=1}^{J} p_{ij0} E\left\{\int_0^{u_i} \frac{\partial\{p_{ij}f_{ij}(t)\}}{p_{ij}f_{ij}(t)\partial\theta} dF_{i0}(t)\right\}_{\theta=\theta_0} \\
&+ E\left\{S_{i0}(u_i)\frac{\partial S_i(u_i)}{S_i(u_i)\partial\theta}\right\}_{\theta=\theta_0} \\
=& \sum_{j=1}^{J} p_{ij0} E\left\{\int_0^{u_i} \frac{\partial\{p_{ij}f_{ij}(t)\}}{p_{ij0}f_{ij0}(t)\partial\theta} dF_{i0}(t)\right\}_{\theta=\theta_0} \\
&+ E\left\{S_{i0}(u_i)\frac{\partial S_i(u_i)}{S_{i0}(u_i)\partial\theta}\right\}_{\theta=\theta_0} \\
=& \sum_{j=1}^{J} E\left\{\int_0^{u_i} \frac{\partial\{p_{ij}f_{ij}(t)\}}{\partial\theta} dt\right\}_{\theta=\theta_0} + E\left\{\frac{\partial S_i(u_i)}{\partial\theta}\right\}_{\theta=\theta_0} \\
=& E\left\{\frac{\partial}{\partial\theta}\left(\int_0^{u_i}\sum_{j=1}^{J} p_{ij}f_{ij}(t)dt + S_i(u_i)\right)\right\}_{\theta=\theta_0} \\
=& E\left\{\frac{\partial}{\partial\theta}\left(F_i(u_i) + S_i(u_i)\right)\right\}_{\theta=\theta_0} \\
=& 0
\end{aligned}
\tag{3.6.52}
$$

for $i = 1, \ldots, n$. As a result,

$$
E\{\mathcal{S}_n(\theta_0)\} = \sum_{i=1}^{n} E\left\{\frac{\partial l_{ni}(\theta)}{\partial\theta}\right\}_{\theta=\theta_0} = 0. \tag{3.6.53}
$$

Similarly, the second derivative of l_{ni} is

$$
\begin{aligned}
\frac{\partial^2 l_{ni}(\theta)}{\partial\theta\partial\theta^T} =& \sum_{j=1}^{J} \delta_{ij}\left\{\frac{\partial^2\{p_{ij}f_{ij}(t_i)\}}{p_{ij}f_{ij}(t_i)\partial\theta\partial\theta^T} - \frac{1}{p_{ij}^2 f_{ij}^2(t_i)}\frac{\partial\{p_{ij}f_{ij}(t_i)\}}{\partial\theta}\frac{\partial\{p_{ij}f_{ij}(t_i)\}}{\partial\theta^T}\right\} \\
&+ (1-\delta_i)\left\{\frac{\partial^2 S_i(t_i)}{S_i(t_i)\partial\theta\partial\theta^T} - \frac{1}{S_i^2(t_i)}\frac{\partial F_i(t_i)}{\partial\theta}\frac{\partial F_i(t_i)}{\partial\theta^T}\right\}
\end{aligned}
$$

for $i = 1, \ldots, n$. Hence

$$
\begin{aligned}
& E\left\{\frac{\partial^2 l_{ni}(\theta)}{\partial\theta\partial\theta^T}\right\}_{\theta=\theta_0} \\
&= \sum_{j=1}^{J} p_{ij0} E\left\{\int_0^{u_i}\left\{\frac{\partial^2\{p_{ij}f_{ij}(t)\}}{p_{ij}f_{ij}(t)\partial\theta\partial\theta^T} - \frac{1}{p_{ij}^2 f_{ij}^2(t)}\frac{\partial\{p_{ij}f_{ij}(t)\}}{\partial\theta}\frac{\partial\{p_{ij}f_{ij}(t)\}}{\partial\theta^T}\right\} dF_{ij0}(t)\right\}_{\theta=\theta_0} \\
&\quad + E\left\{\left\{\frac{\partial^2 S_i(u_i)}{S_i(u_i)\partial\theta\partial\theta^T} - \frac{1}{S_i^2(u_i)}\frac{\partial F_i(u_i)}{\partial\theta}\frac{\partial F_i(u_i)}{\partial\theta^T}\right\} S_{i0}(u_i)\right\}_{\theta=\theta_0} \\
&= \sum_{j=1}^{J} p_{ij0} E\left\{\int_0^{u_i}\left\{\frac{\partial^2\{p_{ij}f_{ij}(t)\}}{p_{ij0}f_{ij0}(t)\partial\theta\partial\theta^T} - \frac{1}{p_{ij0}^2 f_{ij0}^2(t)}\frac{\partial\{p_{ij}f_{ij}(t)\}}{\partial\theta}\frac{\partial\{p_{ij}f_{ij}(t)\}}{\partial\theta^T}\right\} f_{ij0}(t)dt\right\}_{\theta=\theta_0} \\
&\quad + E\left\{\left\{\frac{\partial^2 S_i(u_i)}{S_{i0}(u_i)\partial\theta\partial\theta^T} - \frac{1}{S_{i0}^2(u_i)}\frac{\partial F_i(u_i)}{\partial\theta}\frac{\partial F_i(u_i)}{\partial\theta^T}\right\} S_{i0}(u_i)\right\}_{\theta=\theta_0} \\
&= \sum_{j=1}^{J} E\left\{\int_0^{u_i}\left\{\frac{\partial^2\{p_{ij}f_{ij}(t)\}}{\partial\theta\partial\theta^T} - \frac{1}{p_{ij0}f_{ij0}(t)}\frac{\partial\{p_{ij}f_{ij}(t)\}}{\partial\theta}\frac{\partial\{p_{ij}f_{ij}(t)\}}{\partial\theta^T}\right\} dt\right\}_{\theta=\theta_0} \\
&\quad + E\left\{\left\{\frac{\partial^2 S_i(u_i)}{\partial\theta\partial\theta^T} - \frac{1}{S_{i0}(u_i)}\frac{\partial S_i(u_i)}{\partial\theta}\frac{\partial S_i(u_i)}{\partial\theta^T}\right\}\right\}_{\theta=\theta_0} \\
&= E\left\{\int_0^{u_i}\frac{\sum_{j=1}^{J}\partial^2\{p_{ij}f_{ij}(t)\}}{\partial\theta\partial\theta^T}dt + \frac{\partial^2 S_i(u_i)}{\partial\theta\partial\theta^T}\right\}_{\theta=\theta_0} \\
&\quad - E\left\{\int_0^{u_i}\sum_{j=1}^{J}\frac{1}{p_{ij0}f_{ij0}(t)}\frac{\partial\{p_{ij}f_{ij}(t)\}}{\partial\theta}\frac{\partial\{p_{ij}f_{ij}(t)\}}{\partial\theta^T}dt + \frac{1}{S_{i0}(u_i)}\frac{\partial S_i(u_i)}{\partial\theta}\frac{\partial S_i(u_i)}{\partial\theta^T}\right\} \\
&= -E\left\{\int_0^{u_i}\sum_{j=1}^{J}\frac{1}{p_{ij0}f_{ij0}(t)}\frac{\partial\{p_{ij}f_{ij}(t)\}}{\partial\theta}\frac{\partial\{p_{ij}f_{ij}(t)\}}{\partial\theta^T}dt + \frac{1}{S_{i0}(u_i)}\frac{\partial S_i(u_i)}{\partial\theta}\frac{\partial S_i(u_i)}{\partial\theta^T}\right\}.
\end{aligned}
$$

Note that $\delta_{ij}^2 = \delta_{ij}$, $(1-\delta_i)^2 = 1-\delta_i$ and $\delta_{ij}(1-\delta_i) = 0$ by their definition, we also have

$$
\begin{aligned}
& E\left\{\frac{\partial l_{ni}(\theta)}{\partial\theta}\frac{\partial l_{ni}(\theta)}{\partial\theta^T}\right\}_{\theta=\theta_0} \\
&= E\left\{\sum_{j=1}^{J}\frac{\delta_{ij}}{p_{ij}^2 f_{ij}^2(t_i)}\frac{\partial\{p_{ij}f_{ij}(t_i)\}}{\partial\theta}\frac{\partial\{p_{ij}f_{ij}(t_i)\}}{\partial\theta^T} + \frac{1-\delta_i}{S_i^2(t_i)}\frac{\partial S_i(t_i)}{\partial\theta}\frac{\partial S_i(t_i)}{\partial\theta^T}\right\}_{\theta=\theta_0} \\
&= E\left\{\int_0^{u_i}\sum_{j=1}^{J}\frac{1}{p_{ij0}f_{ij0}(t)}\frac{\partial\{p_{ij}f_{ij}(t_i)\}}{\partial\theta}\frac{\partial\{p_{ij}f_{ij}(t_i)\}}{\partial\theta^T}dt + \frac{1}{S_{i0}(u_i)}\frac{\partial S_i(t_i)}{\partial\theta}\frac{\partial S_i(t_i)}{\partial\theta^T}\right\}_{\theta=\theta_0} \\
&= -E\left\{\frac{\partial^2 l_{ni}(\theta)}{\partial\theta\partial\theta^T}\right\}_{\theta=\theta_0}.
\end{aligned}
$$

As the individuals are assumed to be independent of each other and (3.6.52) and (3.6.53) holds, we get

$$\begin{aligned} E\{\mathcal{S}_n(\theta)(\mathcal{S}_n(\theta))^T\}_{\theta=\theta_0} &= \sum_{i=1}^{n} var(\mathcal{S}_n(\theta))_{\theta=\theta_0} \\ &= \sum_{i=1}^{n} var\left(\frac{\partial l_{in}(\theta)}{\partial\theta}\right)_{\theta=\theta_0} \\ &= \sum_{i=1}^{n} E\left(\frac{\partial l_{in}(\theta)}{\partial\theta}\frac{\partial l_{in}(\theta)}{\partial\theta^T}\right)_{\theta=\theta_0} \\ &= -\sum_{i=1}^{n} E\left\{\frac{\partial^2 l_{ni}(\theta)}{\partial\theta\partial\theta^T}\right\}_{\theta=\theta_0} \\ &= E\{\mathcal{F}_n(\theta_0)\}. \end{aligned} \tag{3.6.54}$$

Hence condition (C1) holds by (3.6.53), (3.6.54) and the regularity conditions (R2)-(R3). ■

Lemma 3.6.2 *Assume that conditions* ($B1$) *holds, then condition* ($C2$) *holds.*

Proof. The results follow readily from the proof of Theorem 2 in Choi and zhou (2002). Thus we omit the proof here. ■

Lemma 3.6.3 *If (B1) and (B2) are satisfied, then as* $n \to \infty$,

$$||\mathcal{Q}_n^{-1/2}\{\mathcal{F}_n(\theta_0) - \mathcal{Q}_n\}(\mathcal{Q}_n^{-1/2})^T||_1 \xrightarrow{P} 0. \tag{3.6.55}$$

Proof. Let v be an arbitrary unit vector in $\mathbb{R}^{\pi_i JK+(\pi_2+1)J}$, thus by routine calculus we have

$$
\begin{aligned}
& v^T\{\mathcal{Q}_n^{-1/2}\{\mathcal{F}_n(\theta_0)-\mathcal{Q}_n\}(\mathcal{Q}_n^{-1/2})^T\}v \\
&=\sum_{i=1}^{n} v^T\{\mathcal{Q}_n^{-1/2}\{H_i\mathscr{F}_i(\theta_0)H_i^T-H_i\mathscr{Q}_i(\theta_0)H_i^T\}(\mathcal{Q}_n^{-1/2})^T\}v \\
&=\sum_{i=1}^{n}\sum_{r=1}^{\pi_i JK+(\pi_2+1)J}\sum_{m=1}^{\pi_i JK+(\pi_2+1)J}\{f_i^{rm}(\theta_0)-q_i^{rm}\}v^T e_{ir}e_{im}^T v, \qquad (3.6.56)
\end{aligned}
$$

where e_{ir} is the rth column vector of $\mathcal{Q}_n^{-1/2}H_i$ for $r=1,\ldots,\pi_i JK+(\pi_2+1)J$ and $i=1,\ldots,n$, i.e.

$$\mathcal{Q}_n^{-1/2}H_i=(e_{i1},\ldots,e_{i,\pi_i JK+(\pi_2+1)J}). \qquad (3.6.57)$$

Furthermore, as $E[f_i^{rm}(\theta_0)]=q_i^{rm}$, we have

$$E\{v^T\{\mathcal{Q}_n^{-1/2}\{\mathcal{F}_n(\theta_0)-\mathcal{Q}_n\}(\mathcal{Q}_n^{-1/2})^T\}v\}=0.$$

Let $\mathcal{W}_{irm}=\{f_i^{rm}(\theta_0)-q_i^{rm}\}v^T e_{ir}e_{im}^T v$, it is obvious that $E[\mathcal{W}_{irm}]=0$, thus

$$var(\mathcal{W}_{irm})=E\{\{f_i^{rm}(\theta_0)-q_i^{rm}\}^2(v^T e_{ir}e_{im}^T v)^2\}=(v^T e_{ir}e_{im}^T v)^2 var\{f_i^{rm}\}.$$

When $(i_1-i_2)^2+(r_1-r_2)^2+(m_1-m_2)^2>0$, we get

$$
\begin{aligned}
cov(\mathcal{W}_{i_1r_1m_1},\mathcal{W}_{i_2r_2m_2}) &= E[\mathcal{W}_{i_1r_1m_1}\mathcal{W}_{i_2r_2m_2}] \\
&\le E\left\{\frac{\mathcal{W}_{i_1r_1m_1}^2+\mathcal{W}_{i_2r_2m_2}^2}{2}\right\} \\
&=\frac{1}{2}\{var(\mathcal{W}_{i_1r_1m_1})\}+\frac{1}{2}\{var(\mathcal{W}_{i_2r_2m_2})\}.
\end{aligned}
$$

As a result of the Cauchy-Schwarz inequality,

$$
\begin{aligned}
&\lim_{n\to\infty} var\{v^T\{\mathcal{Q}_n^{-1/2}\{\mathcal{F}_n(\theta_0)-\mathcal{Q}_n\}(\mathcal{Q}_n^{-1/2})^T\}v\} \\
&\le \{\pi_i JK+(\pi_2+1)J\}^2 \lim_{n\to\infty}\sum_{i=1}^{n}\sum_{r=1}^{\pi_i JK+(\pi_2+1)J}\sum_{m=1}^{\pi_i JK+(\pi_2+1)J} var\{f_i^{rm}(\theta_0)\}(v^T e_{ir}e_{im}^T v)^2. \\
&\le M\{\pi_i JK+(\pi_2+1)J\}^2 \lim_{n\to\infty}\sum_{i=1}^{n}\sum_{r=1}^{\pi_i JK+(\pi_2+1)J}\sum_{m=1}^{\pi_i JK+(\pi_2+1)J} (v^T e_{ir}e_{im}^T v)^2 \\
&\le M\{\pi_i JK+(\pi_2+1)J\}^2 \lim_{n\to\infty}\sum_{i=1}^{n}\sum_{r=1}^{\pi_i JK+(\pi_2+1)J}\sum_{m=1}^{\pi_i JK+(\pi_2+1)J} |e_{ir}|^2|e_{im}|^2 \\
&= M\{\pi_i JK+(\pi_2+1)J\}^2 \lim_{n\to\infty}\sum_{i=1}^{n}(\sum_{r=1}^{\pi_i JK+(\pi_2+1)J} |e_{ir}|^2)^2 \\
&= M\{\pi_i JK+(\pi_2+1)J\}^2 \lim_{n\to\infty}\sum_{i=1}^{n}(tr\{H_i^T\mathcal{Q}_n^{-1}H_i\})^2 \\
&= 0. \qquad (3.6.58)
\end{aligned}
$$

Then (3.6.55) follows from Chebychev's inequality. ∎

Lemma 3.6.4 *If (B1) and (B3) hold, then for each $A>0$, as $n\to\infty$,*

$$\sup_{\theta\in\mathcal{N}_n(A)} ||\mathcal{Q}_n^{-1/2}\{\mathcal{F}_n(\theta)-\mathcal{F}_n(\theta_0)\}(\mathcal{Q}_n^{-1/2})^T||_1 \xrightarrow{P} 0. \qquad (3.6.59)$$

Proof. Let v be an arbitrary unit vector in $\mathbb{R}^{\pi_i JK+(\pi_2+1)J}$ and e_{ir} be defined by (3.6.57), then follows from $(B1)$, $(B3)$ and Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&\lim_{n\to\infty} E\left\{\sup_{\theta\in\mathcal{N}_n(A)} |v^T \mathcal{Q}_n^{-1/2}\{\mathcal{F}_n(\theta)-\mathcal{F}_n(\theta_0)\}(\mathcal{Q}_n^{-1/2})^T v|\right\} \\
&= \lim_{n\to\infty}\sum_{i=1}^{n}\sum_{r=1}^{\pi_i JK+(\pi_2+1)J}\sum_{m=1}^{\pi_i JK+(\pi_2+1)J} E\left\{\sup_{\theta\in\mathcal{N}_n(A)} |f_i^{rm}(\theta)-f_i^{rm}(\theta_0)||v^T e_{ir}e_{im}^T v|\right\} \\
&\le M^{'}A\lim_{n\to\infty}\sum_{i=1}^{n}\sum_{r=1}^{\pi_i JK+(\pi_2+1)J}\sum_{m=1}^{\pi_i JK+(\pi_2+1)J} (tr\{H_i^T\mathcal{Q}_n^{-1}H_i\})^{1/2}|e_{ir}||e_{im}| \\
&= M^{'}A\lim_{n\to\infty}\sum_{i=1}^{n}(tr\{H_i^T\mathcal{Q}_n^{-1}H_i\})^{1/2}\left\{\sum_{m=1}^{\pi_i JK+(\pi_2+1)J}|e_{im}|\right\}^2 \\
&\le M^{'}A(\pi_i JK+(\pi_2+1)J)\lim_{n\to\infty}\sum_{i=1}^{n}(tr\{H_i^T\mathcal{Q}_n^{-1}H_i\})^{1/2}\left\{\sum_{m=1}^{\pi_i JK+(\pi_2+1)J}|e_{im}|^2\right\} \\
&\le M^{'}A(\pi_i JK+(\pi_2+1)J)\lim_{n\to\infty}\sum_{i=1}^{n}(tr\{H_i^T\mathcal{Q}_n^{-1}H_i\})^{3/2} \\
&= 0.
\end{aligned}$$

Hence (3.6.55) is implied by Markov's inequality. ∎

Lemma 3.6.5 *If $(B1)$, $(B2)$ and $(B3)$ are satisfied, then $(C3)$ holds.*

Proof. Note that

$$\begin{aligned}
&\mathcal{Q}_n^{-1/2}\{\mathcal{F}_n(\theta)\}(\mathcal{Q}_n^{-1/2})^T \\
&= \mathcal{Q}_n^{-1/2}\{\mathcal{F}_n(\theta)-\mathcal{F}_n(\theta_0)+\mathcal{F}_n(\theta_0)-\mathcal{Q}_n+\mathcal{Q}_n\}(\mathcal{Q}_n^{-1/2})^T \\
&= \mathcal{Q}_n^{-1/2}\{\mathcal{F}_n(\theta)-\mathcal{F}_n(\theta_0)\}(\mathcal{Q}_n^{-1/2})^T+\mathcal{Q}_n^{-1/2}\{\mathcal{F}_n(\theta_0)-\mathcal{Q}_n\}(\mathcal{Q}_n^{-1/2})^T+\mathcal{Q}_n^{-1/2}\{\mathcal{Q}_n\}(\mathcal{Q}_n^{-1/2})^T \\
&= I+\mathcal{Q}_n^{-1/2}\{\mathcal{F}_n(\theta_0)-\mathcal{Q}_n\}(\mathcal{Q}_n^{-1/2})^T+\mathcal{Q}_n^{-1/2}\{\mathcal{F}_n(\theta)-\mathcal{F}_n(\theta_0)\}(\mathcal{Q}_n^{-1/2})^T. \qquad (3.6.60)
\end{aligned}$$

The right side of (3.6.60) converges in probability towards I by (3.6.55) and (3.6.59) and so does the left side. $(C3)$ holds consequently. ∎

Lemma 3.6.6 *If* $(B1)$ *and* $(B4)$ *are satisfied, then* $(C4)$ *holds.*

Proof. Note that the individuals under consideration are independent of each other. This proof will appeal to the central limit theorem. Let v be an arbitrary unit vector, thus

$$v^T \mathcal{Q}_n^{-1/2} \mathcal{S}_n(\theta_0) = \sum_{i=1}^{n} v^T \mathcal{Q}_n^{-1/2} H_i \mathscr{S}_i(\theta_0). \tag{3.6.61}$$

By (3.6.52), we can get

$$E\{v^T \mathcal{Q}_n^{-1/2} \mathcal{S}_n(\theta_0)\} = E\left\{v^T \mathcal{Q}_n^{-1/2} H_i \mathscr{S}_i(\theta_0)\right\} = 0.$$

We still have

$$\begin{aligned} var\{v^T \mathcal{Q}_n^{-1/2} \mathcal{S}_n(\theta_0)\} &= E\left\{v^T \mathcal{Q}_n^{-1/2} \mathcal{S}_n(\theta_0) \mathcal{S}_n(\theta_0)^T (\mathcal{Q}_n^{-1/2})^T v\right\} \\ &= v^T \mathcal{Q}_n^{-1/2} \mathcal{Q}_n (\mathcal{Q}_n^{-1/2})^T v \\ &= 1. \end{aligned}$$

By condition $(C1)$ and Cauchy-Schwarz inequalities, we have

$$\begin{aligned} E\left\{v^T \mathcal{Q}_n^{-1/2} H_i \mathscr{S}_i(\theta_0)\right\}^2 &= E\left\{v^T \mathcal{Q}_n^{-1/2} H_i \mathscr{S}_i(\theta_0) \mathscr{S}_i(\theta_0)^T H_i^T (\mathcal{Q}_n^{-1/2})^T v\right\} \\ &= E\{\mathscr{S}_i(\theta_0) \mathscr{S}_i(\theta_0)^T\} \left\{v^T \mathcal{Q}_n^{-1/2} H_i H_i^T (\mathcal{Q}_n^{-1/2})^T v\right\} \\ &\le E|\mathscr{S}_i(\theta_0)|^2 \lambda_{max} \left\{v^T \mathcal{Q}_n^{-1/2} H_i H_i^T (\mathcal{Q}_n^{-1/2})^T v\right\} \\ &\le E|\mathscr{S}_i(\theta_0)|^2 tr \left\{v^T \mathcal{Q}_n^{-1/2} H_i H_i^T (\mathcal{Q}_n^{-1/2})^T v\right\} \end{aligned} \tag{3.6.62}$$

and similarly,

$$
\begin{aligned}
E\left\{v^T \mathcal{Q}_n^{-1/2} H_i \mathscr{S}_i(\theta_0)\right\}^4 &= E\left\{v^T \mathcal{Q}_n^{-1/2} H_i \mathscr{S}_i(\theta_0) \mathscr{S}_i(\theta_0)^T H_i^T (\mathcal{Q}_n^{-1/2})^T v\right\}^2 \\
&= E\{\mathscr{S}_i(\theta_0)\mathscr{S}_i(\theta_0)^T\}^2 \left\{v^T \mathcal{Q}_n^{-1/2} H_i H_i^T (\mathcal{Q}_n^{-1/2})^T v\right\}^2 \\
&\leq E|\mathscr{S}_i(\theta_0)|^4 \lambda_{max} \left\{v^T \mathcal{Q}_n^{-1/2} H_i H_i^T (\mathcal{Q}_n^{-1/2})^T v\right\}^2 \\
&\leq E|\mathscr{S}_i(\theta_0)|^4 tr \left\{v^T \mathcal{Q}_n^{-1/2} H_i H_i^T (\mathcal{Q}_n^{-1/2})^T v\right\}^2 .
\end{aligned}
\tag{3.6.63}
$$

Note that Cauchy-Schwarz inequalities in addition with condition $(B4)$ imply that

$$
\begin{aligned}
E|\mathscr{S}_i(\theta_0)|^4 &= E\left\{\sum_{r=1}^{\pi_i JK + (\pi_2+1)J} [s_i^r(\theta_0)]^2\right\}^2 \\
&\leq \{\pi_i JK + (\pi_2+1)J\} \left\{\sum_{r=1}^{\pi_i JK + (\pi_2+1)J} E[s_i^r(\theta_0)]^4\right\} \\
&\leq \{\pi_i JK + (\pi_2+1)J\}^2 M''
\end{aligned}
$$

and that

$$
\begin{aligned}
E|\mathscr{S}_i(\theta_0)|^2 &= \int |\mathscr{S}_i(\theta_0)|^2 d\mathbb{P} \\
&\leq \left\{\int |\mathscr{S}_i(\theta_0)|^4 d\mathbb{P}\right\}^{\frac{1}{2}} \\
&= \left\{E|\mathscr{S}_i(\theta_0)|^4\right\}^{\frac{1}{2}} \\
&\leq \sqrt{M''},
\end{aligned}
$$

where $\mathbb{P}$ is the probability measure with which the expectation is respect to. As a result, we have

$$
E\left\{v^T \mathcal{Q}_n^{-1/2} H_i \mathscr{S}_i(\theta_0)\right\}^2 \leq \sqrt{M''} tr \left\{v^T \mathcal{Q}_n^{-1/2} H_i H_i^T (\mathcal{Q}_n^{-1/2})^T v\right\} \tag{3.6.64}
$$

and

$$E\left\{v^T\mathcal{Q}_n^{-1/2}H_i\mathscr{S}_i(\theta_0)\right\}^4 \leq \{\pi_i JK + (\pi_2+1)J\}^2 M'' tr\left\{v^T\mathcal{Q}_n^{-1/2}H_iH_i^T(\mathcal{Q}_n^{-1/2})^T v\right\}^2.$$

Thus by condition $(B1)$, Cauchy-Schwarz inequalities and Chebyshev's inequalities, we have, for any $\epsilon > 0$,

$$\begin{aligned}
0 &\leq \lim_{n\to\infty}\sum_{i=1}^{n}\int_{|v^T\mathcal{Q}_n^{-1/2}H_i\mathscr{S}_i(\theta_0)|\geq\epsilon}\{v^T\mathcal{Q}_n^{-1/2}H_i\mathscr{S}_i(\theta_0)\}^2 d\mathbb{P} \\
&= \lim_{n\to\infty}\sum_{i=1}^{n}\int 1_{\{|v^T\mathcal{Q}_n^{-1/2}H_i\mathscr{S}_i(\theta_0)|\geq\epsilon\}}\{v^T\mathcal{Q}_n^{-1/2}H_i\mathscr{S}_i(\theta_0)\}^2 d\mathbb{P} \\
&\leq \lim_{n\to\infty}\sum_{i=1}^{n}\left\{\int 1_{\{|v^T\mathcal{Q}_n^{-1/2}H_i\mathscr{S}_i(\theta_0)|\geq\epsilon\}}d\mathbb{P}\right\}^{\frac{1}{2}}\left\{\int\{v^T\mathcal{Q}_n^{-1/2}H_i\mathscr{S}_i(\theta_0)\}^4 d\mathbb{P}\right\}^{\frac{1}{2}} \\
&= \lim_{n\to\infty}\sum_{i=1}^{n}P^{1/2}\{|v^T\mathcal{Q}_n^{-1/2}H_i\mathscr{S}_i(\theta_0)|\geq\epsilon\}\left\{\int\{v^T\mathcal{Q}_n^{-1/2}H_i\mathscr{S}_i(\theta_0)\}^4 d\mathbb{P}\right\}^{\frac{1}{2}} \\
&\leq \lim_{n\to\infty}\sum_{i=1}^{n}\left\{\frac{var\{v^T\mathcal{Q}_n^{-1/2}H_i\mathscr{S}_i(\theta_0)\}}{\epsilon^2}\right\}^{\frac{1}{2}}\left\{\int\{v^T\mathcal{Q}_n^{-1/2}H_i\mathscr{S}_i(\theta_0)\}^4 d\mathbb{P}\right\}^{\frac{1}{2}} \\
&= \lim_{n\to\infty}\frac{1}{\epsilon}\sum_{i=1}^{n}\left\{E\{v^T\mathcal{Q}_n^{-1/2}H_i\mathscr{S}_i(\theta_0)\}^2\right\}^{\frac{1}{2}}\left\{E\{v^T\mathcal{Q}_n^{-1/2}H_i\mathscr{S}_i(\theta_0)\}^4\right\}^{\frac{1}{2}} \\
&\leq \lim_{n\to\infty}\frac{1}{\epsilon}\sum_{i=1}^{n}M''^{\frac{3}{4}}\left\{tr\left\{v^T\mathcal{Q}_n^{-1/2}H_iH_i^T(\mathcal{Q}_n^{-1/2})^T v\right\}\right\}^{\frac{3}{2}} \\
&= 0.
\end{aligned}$$

As a result,

$$\lim_{n\to\infty}\sum_{i=1}^{n}\int_{|v^T\mathcal{Q}_n^{-1/2}H_i\mathscr{S}_i(\theta_0)|\geq\epsilon}\{v^T\mathcal{Q}_n^{-1/2}H_i\mathscr{S}_i(\theta_0)\}^2 d\mathbb{P}.$$

The Lindeberg condition holds consequently. So as $n\to\infty$

$$v^T\mathcal{Q}_n^{-1/2}\mathcal{S}_n(\theta_0)\xrightarrow{D}N(0,1),$$

where $N(0,1)$ is a standard normal distribution. As v is an arbitrary unit vector, the elements of $\mathcal{Q}_n^{-1/2}\mathcal{S}_n(\theta_0)$ converge in distribution to $N(0,1)$. Note that the individuals are independent of each other, condition $(C4)$ holds consequently. ■

Proof of Theorem 3.3.1. Since conditions $(C1)-(C4)$ have been verified by the lemmas, it suffices to check the conditions $(A1)-(A2)$ to use Theorem 2.1 in Vu and Zhou (1997). The subset over which the maximization takes place is specified as:

$$\tau = \Theta_1 = \mathbb{R}^{\pi_1 JK + (\pi_2+1)J}.$$

On the basis of the specification, we can recast conditions $(A1)-(A3)$ as

(D1) For a neighborhood $\mathscr{N}$ of θ_0, the function $l_n(\theta)$ is continuous on $\tau \cap \mathcal{N}$, and the first and second directional derivatives of $l_n(\theta)$ with respect to θ exist, are finite and are continuous on $\tau \cap \mathcal{N}$.

(D2) There is a closed cone C_τ with vertex at θ_0 such that

$$C_\tau \subseteq \Theta_1 \qquad \text{and} \qquad C_\tau \cap \mathcal{N} = \tau \cap \mathcal{N},$$

where $\mathcal{N}$ is a closed neighborhood of θ_0.

Condition $(D1)$ is satisfied trivially by the regularity conditions. Let

$$C_\tau = \tau = \mathbb{R}^{\pi_1 JK + (\pi_2+1)J}.$$

It is clear that $\mathcal{N}(\delta) \subseteq \tau$ for sufficiently small δ, since θ_0 is an interior point of Θ_1. Then condition (D2) is established.

Now conditions $(C1)-(C4)$ and $(D1)-(D2)$ are verified. Then Theorem 2.1 of Vu and Zhou (1997) gives a local maximum likelihood estimator $\hat{\theta}_n^{(1)}$ with maximization taking place over τ, which is locally uniquely determined interior to θ_0 with probability approaching 1, and is consistent for θ_0 as $n \to \infty$. The asymptotic distributions of the maximum likelihood estimators, however, are not discussed in Vu and Zhou (1997). The next part of the proof is for the asymptotic normality of $\hat{\theta}_n^{(1)}$. Let v be an arbitrary unit vector in $\mathbb{R}^{\pi_i JK+(\pi_2+1)J}$. Note that $\hat{\theta}_n^{(1)}$ is a maximizer of $l_n(\theta)$ over a neighborhood $\mathcal{N}_n(A)$ of θ_0 with probability approaching 1, i.e.

$$\mathcal{S}_n(\hat{\theta}_n^{(1)}) = 0 \tag{3.6.65}$$

with probability approaching 1. Let θ^* be on the line segment between $\hat{\theta}_n^{(1)}$ and θ_0, then by Taylor expansion we have

$$\begin{aligned} v^T \mathcal{Q}_n^{-1/2} \mathcal{S}_n(\theta_0) &= v^T \mathcal{Q}_n^{-1/2} \mathcal{S}_n(\theta_0) - v^T \mathcal{Q}_n^{-1/2} \mathcal{S}_n(\hat{\theta}_n^{(1)}) \\ &= v^T \mathcal{Q}_n^{-1/2} \mathcal{F}_n(\theta^*)(\hat{\theta}_n^{(1)} - \theta_0) + o_p(1) \\ &= v^T \left\{ I + \mathcal{Q}_n^{-1/2} \{\mathcal{F}_n(\theta^*) - \mathcal{Q}_n\} (\mathcal{Q}_n^{-1/2})^T \right\} (\mathcal{Q}_n^{1/2})^T (\hat{\theta}_n^{(1)} - \theta_0) + o_p(1) \\ &= v^T \left\{ I + o_p(1) \right\} (\mathcal{Q}_n^{1/2})^T (\hat{\theta}_n^{(1)} - \theta_0) + o_p(1), \end{aligned} \tag{3.6.66}$$

where $o_p(1)$ is a square matrix whose elements converge in probability to 0. Hence $v^T \mathcal{Q}_n^{-1/2} \mathcal{S}_n(\theta_0)$ and $v^T (\mathcal{Q}_n^{1/2})^T (\hat{\theta}_n^{(1)} - \theta_0)$ have the same limiting distribution. Thus (3.3.37) holds. To prove (3.3.38), it suffices to show that as $n \to \infty$,

$$(\mathcal{Q}_n^{1/2})^T (\hat{\theta}_n^{(1)})(\hat{\theta}_n^{(1)} - \theta_0) - (\mathcal{F}_n^{1/2})^T (\hat{\theta}_n^{(1)})(\hat{\theta}_n^{(1)} - \theta_0) \xrightarrow{D} 0. \tag{3.6.67}$$

It holds directly from condition $(C3)$.

Proof of Theorem 3.3.2. **and Theorem** 3.3.3. For these theorems, we shall use the results of Vu and Zhou (1997) again. Here we need to deal with two parameter spaces τ_1 and τ_2, where τ_1 is a subset of Θ_1 to which the true point θ_0 is constrained under the null hypothesis and τ_2 is the parameter space under the alternative hypothesis. The requirements for τ in conditions $(D1)$ and $(D2)$ should be satisfied by τ_1 and τ_2. In addition, τ_1 and τ_2 are required to have similar properties as $(A3)$ in Vu and Zhou (1997) which we recast as condition $(D3)$ here:

(D3) The space τ is said to satisfy $(D3)$ if there exists a closed cone C_τ with vertex at the true point θ_0 and a closed cone $C_\tau^{'}$ with vertex at 0 such that C_τ can be rescaled to $C_\tau^{'}$.

Let

$$\tau_1 = S_r + \theta^* \subset \Theta_1$$

with $S_r + \theta^*$ being defined by (3.3.39). The parameter space under the unrestricted alternative hypothesis is simply

$$\tau_2 = \Theta_1 = \mathbb{R}^{\pi_1 JK + (\pi_2+1)J}.$$

Further assume that

$$C_{\tau_1} = S_r + \theta^* \qquad \text{and} \qquad C_{\tau_1}^{'} = \mathbb{R}^r \times \{0\}^{\pi_1 JK + (\pi_2+1)J - r}.$$

As θ_0 and θ^* are in the interior of Θ_1, $S_r + \theta^*$ can be rescaled to $\mathbb{R}^r \times \{0\}^{\pi_1 JK + (\pi_2+1)J - r}$ and condition $(D3)$ is satisfied by τ_1. Let

$$C_{\tau_2} = \mathbb{R}^{\pi_1 JK + (\pi_2+1)J} \qquad \text{and} \qquad C_{\tau_2}^{'} = \mathbb{R}^{\pi_1 JK + (\pi_2+1)J}.$$

Condition $(D3)$ holds trivially for τ_2 since $\mathbb{R}^{\pi_1 JK + (\pi_2+1)J}$ keeps invariant under rescaling and centering. Thus the existence, uniqueness and consistency of $\hat{\theta}_n^{(2)}$, the maximum likelihood estimator under the null hypothesis, are implied by Theorem 2.1 of Vu and Zhou(1997) and Theorem 3.3.2 is proved. By Theorem 2.2 of Vu and Zhou(1997) we can get

$$
\begin{aligned}
d_n^{(11)} &= -2(l_n(\hat{\theta}_n^{(2)}) - \hat{\theta}_n^{(1)}) \\
&\xrightarrow{D} \inf_{\theta \in C'_{\tau_1}} |N_{\pi_1 JK+\pi_2 J+J} - \theta| - \inf_{\theta \in C'_{\tau_2}} |N_{\pi_1 JK+\pi_2 J+J} - \theta| \\
&= \inf_{\theta \in \mathbb{R}^r} |N_{\pi_1 JK+\pi_2 J+J} - \theta| - \inf_{\theta \in \mathbb{R}^{\pi_1 JK+\pi_2 J+J}} |N_{\pi_1 JK+\pi_2 J+J} - \theta| \\
&= \chi^2_{\pi_1 JK+\pi_2 J+J-r},
\end{aligned}
$$

where $N_{\pi_1 JK+\pi_2 J+J}$ is a standard normal random vector in $\pi_1 JK+\pi_2 J+J$ dimensions. This proves Theorem 3.3.3.

Proof of Theorem 3.3.4 **and Theorem** 3.3.5**.** Theorem 3.3.4 and Theorem 3.3.5 are for the boundary cases. The proof is similar to that of Theorem 2. The parameter spaces under null hypothesis and alternative hypothesis are respectively

$$\Theta_3 = \mathbb{R}^{\pi_1 JK+(\pi_2+1)(J-1)} \times \{0\}$$

and

$$\Theta_4 = \mathbb{R}^{\pi_1 JK+(\pi_2+1)(J-1)} \times [0, \infty).$$

Further we specify

$$C_{\Theta_3} = \mathbb{R}^{\pi_1 JK+(\pi_2+1)(J-1)} \times \{0\}, \qquad C'_{\Theta_3} = \mathbb{R}^{\pi_1 JK+(\pi_2+1)(J-1)} \times \{0\} \tag{3.6.68}$$

and

$$C_{\Theta_4} = \mathbb{R}^{\pi_1 JK+(\pi_2+1)(J-1)} \times [0, \infty), \qquad C'_{\Theta_4} = \mathbb{R}^{\pi_1 JK+(\pi_2+1)(J-1)} \times [0, \infty). \tag{3.6.69}$$

To verify the conditions $(D1)-(D3)$, we need to define the neighborhood of θ_0 on the boundary by

$$\begin{aligned}\mathscr{N}(\delta) = &\prod_{j=1}^{J}\prod_{k=1}^{K}\prod_{m=1}^{\pi_1}[\alpha_{jkm}-\delta,\ \alpha_{jkm}+\delta]\\ &\times\prod_{j=1}^{J-1}[\xi_j-\delta,\ \xi_j+\delta]\times\prod_{j=1}^{J}\prod_{m=1}^{\pi_2}[\beta_{jm}-\delta,\ \beta_{jm}+\delta]\times[-\delta,\ \delta].\end{aligned}$$

Thus

$$\begin{aligned}\mathscr{N}(\delta)\cap\Theta_3 = &\prod_{j=1}^{J}\prod_{k=1}^{K}\prod_{m=1}^{\pi_1}[\alpha_{jkm}-\delta,\ \alpha_{jkm}+\delta]\\ &\times\prod_{j=1}^{J-1}[\xi_j-\delta,\ \xi_j+\delta]\times\prod_{j=1}^{J}\prod_{m=1}^{\pi_2}[\beta_{jm}-\delta,\ \beta_{jm}+\delta]\times\{0\}\end{aligned}$$

and

$$\begin{aligned}\mathscr{N}(\delta)\cap\Theta_4 = &\prod_{j=1}^{J}\prod_{k=1}^{K}\prod_{m=1}^{\pi_1}[\alpha_{jkm}-\delta,\ \alpha_{jkm}+\delta]\\ &\times\prod_{j=1}^{J-1}[\xi_j-\delta,\ \xi_j+\delta]\times\prod_{j=1}^{J}\prod_{m=1}^{\pi_2}[\beta_{jm}-\delta,\ \beta_{jm}+\delta]\times[0,\infty).\end{aligned}$$

We still need to define the first and second order derivatives of $l_n(\theta)$ in $\mathscr{N}(\delta)\cap\Theta_3$ and $\mathscr{N}(\delta)\cap\Theta_4$. For those points that are in the interior of $l_n(\theta)$ in $\mathscr{N}(\delta)\cap\Theta_3$ and $\mathscr{N}(\delta)\cap\Theta_4$, the derivatives are set to be the usual directional derivatives. For the points that lie on the boundary, i.e. the points in $\mathscr{N}(\delta)\cap\Theta_3$, let

$$\frac{\partial g(\theta)}{\partial\gamma} = \lim_{h^2\to 0}\frac{g\{\theta+\{0\}^{\pi_1 JK+(\pi_2+1)(J-1)}\times\{h^2\}\}-g(\theta)}{h^2},$$

where $g(\cdot)$ is any function satisfying the regularity conditions. It is obvious that $(D1)$ and $(D2)$ are satisfied by Θ_3 and Θ_4. By Theorem 2.1 of Vu and Zhou (1997), there are maximizers $\hat{\theta}_n^{(3)}$ and $\hat{\theta}_n^{(4)}$ of $l_n(\theta)$ over Θ_3 and Θ_4 respectively, which are locally unique with probability approaching 1, and is consistent for θ_0. Theorem 3.3.4 is proved.

As $\mathbb{R}^{\pi_1 JK+(\pi_2+1)(J-1)}$ is trivially invariant under rescaling and centering, C_{Θ_3} and C_{Θ_4} satisfy condition $(D3)$ by (3.6.68) and (3.6.69). Then by Theorem 2.2 of Vu and Zhou(1997),

$$\begin{aligned}
d_n^{(12)} &= -2(l_n(\hat{\theta}_n^{(3)}) - \hat{\theta}_n^{(4)}) \\
&\xrightarrow{D} \inf_{\theta\in C'_{\Theta_3}} |N_{\pi_1 JK+\pi_2(J-1)+J} - \theta|^2 - \inf_{\theta\in C'_{\Theta_4}} |N_{\pi_1 JK+\pi_2(J-1)+J} - \theta|^2 \\
&= \inf_{\theta\in\mathbb{R}^{\pi_1 JK+(\pi_2+1)(J-1)}\times\{0\}} |N_{\pi_1 JK+\pi_2(J-1)+J} - \theta|^2 \\
&\quad - \inf_{\theta\in\mathbb{R}^{\pi_1 JK+(\pi_2+1)(J-1)}\times[0,\infty)} |N_{\pi_1 JK+\pi_2(J-1)+J} - \theta|^2 \\
&= N^2 - N^2 I_{\{N<0\}} \\
&= N^2 I_{\{N\geq 0\}}, \qquad (3.6.70)
\end{aligned}$$

where $N_{\pi_1 JK+\pi_2(J-1)+J}$ represents a standard normal distribution of $\pi_1 JK + \pi_2(J-1) + J$ dimensions and N is the standard normal distribution. Theorem 3.3.5 is proved consequently.

Chapter 4

A Piecewise Exponential Mixture Model

4.1 Introduction

In this chapter, a piecewise exponential mixture model which is a special case of the general mixture model in Chapter 3 is investigated for competing risks data with covariates and possible immunes. Section 4.2 is devoted to the model development. Section 4.3 is for the inference methods and the large sample properties of the estimators and test statistics. In Section 4.4, a set of real life data is fitted by our model. Section 4.5 is for the concluding remarks. Finally, the proofs are put in Section 4.6.

4.2 Model Specification

A simple and flexible distribution for modeling time to event data is piecewise exponential distribution, which has been widely used because it allows us to approximate almost any baseline hazard function reasonably when the cutpoints of the time duration are specified appropriately. There have been many discussions in the literature in which the applications of piecewise exponential distribution are investigated. Among them Friedman (1982) studied piecewise exponential models for survival data with covariates and Larson and Dinse (1985) approximated the Stanford Heart Transplant data by a boundary piecewise exponential mixture model.

In a piecewise exponential mixture distribution setting, the distribution of the survival time of individual i given that $D_i = j$ is assumed to be

$$f_{ij}(t) = \lambda_{ij}(t) \exp\{-\lambda_{ij}(t)t\}, \qquad 1 \leq i \leq n,\, 1 \leq j \leq J \tag{4.2.1}$$

for $t > 0$, where $\lambda_{ij}(t)$'s are linked to observed covariates x_i's, the sub-vector of the covariates of individual i, by

$$\lambda_{ij}(t) = \exp\{\varsigma_j(t) + \alpha_j^T x_i\}, \qquad 1 \leq i \leq n,\, 1 \leq j \leq J. \tag{4.2.2}$$

The time duration $[0, \infty)$ under our study is partitioned into M exhaustive and mutually exclusive intervals $\mathcal{I}_m = [\tau_{m-1},\, \tau_m)$ with cut points

$$0 = \tau_0 < \tau_1 < \cdots < \tau_M = \infty.$$

In addition, the baseline hazard function is assumed to be constant within each interval, so that

$$\varsigma_j(t) = \varsigma_{jm}, \qquad t \in \mathcal{I}_m,\, 1 \leq m \leq M. \tag{4.2.3}$$

The failure rate with respect to a specific risk j of individual i is related to y_i, another sub-vector of the covariates of individual i, by (3.2.15) for the interior cases and by (3.2.16) and (3.2.17) for boundary cases, i.e.

$$p_{ij} = \frac{\exp(\xi_j + \beta_j^T y_i)}{1 + \sum_{l=1}^{J} \exp(\xi_l + \beta_l^T y_i)} \tag{4.2.4}$$

when the true point lies in the interior of the parameter space and

$$p_{ij} = \frac{\exp(\xi_j + \beta_j^T y_i)}{\gamma + 1 + \sum_{l=1}^{J-1} \exp(\xi_l + \beta_l^T y_i)}, \quad 1 \le j \le J-1 \tag{4.2.5}$$

and

$$p_{iJ} = \frac{1}{\gamma + 1 + \sum_{l=1}^{J-1} \exp(\xi_l + \beta_l^T y_i)} \tag{4.2.6}$$

when the true point lies on the boundary of the parameter space. Under the previous assumptions,

$$\begin{aligned} F_{ij}(t) &= P\{T_i^* \le t | D_i = j\} \\ &= \sum_{m=1}^{M} 1_{\{t > \tau_{m-1}\}} P\{\tau_{m-1} < T_i^* \le (t \wedge \tau_m) | D_i = j\} \\ &= \sum_{m=1}^{M} 1_{\{t > \tau_{m-1}\}} \Big\{ \exp\big(-\tau_{m-1} e^{\varsigma_{jm} + \alpha_j^T x_i}\big) - \exp\big(-(t \wedge \tau_m) \exp^{\varsigma_{jm} + \alpha_j^T x_i}\big) \Big\} \end{aligned}$$

for $1 \le i \le n$ and $1 \le j \le J$. Noting that for any $t < \infty$,

$$P\{T_i^* \le t | D_i = 0\} = 0 \quad \text{and} \quad P\{T_i^* > t | D_i = 0\} = 1,$$

the distribution function of individual i is

$$F_i(t) = P\{T_i^* \le t\} = \sum_{j=1}^{J} p_{ij} F_{ij}(t). \tag{4.2.7}$$

Recall that the observations normally consist of $(t_i, \delta_{ij}, x_i, y_i)$, $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, J$, where t_i represents the observed survival time for the ith individual, x_i

and y_i are the observed values of the covariates and δ_{ij} is the risk indicator defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if individual } i \text{ dies from cause } j, \\ 0 & \text{otherwise,} \end{cases} \qquad i = 1, 2, \ldots, n; \quad j = 1, \ldots, J. \quad (4.2.8)$$

The ith individual who experiences a type j failure contributes $p_{ij} f_{ij}(t_i)$ to the likelihood function. Alternatively, individual i who is censored contributes $1 - F_i(t_i) =: S_i(t_i)$ to the likelihood. As a consequence, the full likelihood function is

$$L_n = \prod_{i=1}^{n} \left\{ \prod_{j=1}^{J} \big(p_{ij} f_{ij}(t_i)\big)^{\delta_{ij}} \big(1 - F_i(t_i)\big)^{1-\delta_i} \right\}.$$

The log-likelihood function of the observations is

$$l_n = \sum_{i=1}^{n} \left\{ \sum_{j=1}^{J} \delta_{ij} \big(\log p_{ij} + \log f_{ij}(t_i) \big) + (1 - \delta_i) \log S_i(t_i) \right\},$$

which is the estimating function in this part of study.

Assume $\varsigma_j = (\varsigma_{j1}, \ldots, \varsigma_{jM})^T$, it should be emphasized that the parameters to be estimated for the interior case are

$$\theta = (\varsigma_1^T, \ldots, \varsigma_J^T, \alpha_1^T, \ldots, \alpha_J^T, \beta_1^T, \ldots, \beta_J^T, \xi_1, \ldots, \xi_J)^T$$

and the parameters to be estimated for the boundary case are

$$\theta = (\varsigma_1^T, \ldots, \varsigma_J^T, \alpha_1^T, \ldots, \alpha_J^T, \beta_1^T, \ldots, \beta_{J-1}^T, \xi_1, \ldots, \xi_{J-1}, \gamma)^T.$$

The estimators of the parameters could be developed under some further specifications. In next section, we will investigate the application of the EM algorithm to the statistical inferences on the mixture piecewise exponential model and develop several tests for hypotheses of practical interest.

4.3 Main Results and Applications

For the piecewise exponential mixture model specified in Section 4.2 with finite follow up, the sufficient conditions $(B2) - (B4)$ can be replaced by the following simpler condition:

(E) The true value of the regression coefficients of the hazard functions are uniformly bounded. More specifically, there is a constant $M_s > 0$ such that

$$|\varsigma_{jm0}| \leq M_s \qquad \text{and} \qquad |\alpha_{j0}| \leq M_s$$

for $1 \leq j \leq J$ and $1 \leq m \leq M$.

Remark. Condition (E) implies that for an individual, the failure rates subject to different types of risks are comparable, i.e. no particular type of risk dominates the other types. If condition (E) is violated, i.e. there is a risk which is negligible compared with another, degeneracy would arise. Furthermore, Condition (E) implies that if the covariates are finite, there exists $0 < \mathcal{E} < 1$ such that

$$\mathcal{E} \leq \lambda_{ij}(t) \leq \frac{1}{\mathcal{E}}$$

for $1 \leq i \leq n$, $\leq j \leq J$ and $0 \leq t < \infty$.

Thus we have the following theorem.

Theorem 4.3.1 *For the piecewise exponential mixture model specified in Section 4.2 with bounded censorship and covariates, if (B1) and (E) are satisfied, the conclusions of Theorem* $3.3.1 - 3.3.5$ *hold.*

Theorem 4.3.1 provides us with the existence, uniqueness and normality with probability approaching 1 of the maximum likelihood estimators of the parameters local to the true point over several subspaces of the parameter space. It also allows us to carry out a number of tests of practical interest by restricting the parameter in the underlying subspaces which depend on the details of the tests.

To calculate the estimates, we adopt EM algorithm which is proposed by Dempster, Laird and Rubin (1977). EM algorithm has been a popular method in statistical computation as the computation based on the likelihood function of complete data is much easier than that on the basis of incomplete data. The convergence properties of EM algorithm was then discussed in Wu (1983). Our settings satisfy the requirement of Theorem 2 in Wu (1983) trivially so that EM algorithm is applicable and a stationary point could be reached. For other examples of applications of EM algorithm in survival analysis, see Larson and Dinse (1985), Peng and Dear (2000) and Craiu and Duchesne (2004).

The EM algorithm is applicable when there is a many-to-one mapping from the complete data to incomplete data. In our analysis, the observations are incomplete because of the existence of censoring. Take individual i for example, the missing information that is unobservable is the failure cause that individual i may be susceptible to and will eventually fail from if the observation of individual i is censored.

A set of "pseudo-data" should be created to use the EM algorithm. Let c_{ij} be a

Bernoulli variable indicating failure causes, i.e.

$$c_{ij} = \begin{cases} 1 & \text{if individual } i \text{ will fail from cause } j \text{ eventually,} \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2, \ldots, n$ and $j = 1, \ldots, J$. Naturally $c_i = \sum_{j=1}^{J} c_{ij}$ can be used to represent the cured status of individual i as

$$c_i = \begin{cases} 1 & \text{if patient } i \text{ is susceptible to the risks,} \\ 0 & \text{if } i \text{ is cured or an immune subject} \end{cases}$$

for $i = 1, 2, \ldots, n$. The relationships of the c_{ij}'s and δ_{ij}'s are

$$c_{ij} = \begin{cases} 1 & \delta_{ij} = 1, \\ 0 & \delta_{ij} = 0 \text{ and } \delta_i = 1, \\ 0 \text{ or } 1 & \delta_i = 0 \end{cases}$$

and

$$\delta_{ij} = \begin{cases} 0 & c_{ij} = 0\,, \\ 0 \text{ or } 1 & c_{ij} = 1 \end{cases}$$

for $i = 1, 2, \ldots, n$ and $j = 1, \ldots, J$.

The complete observation of individual i is assumed to be the combination of c_{ij} and the normal observation $(t_i, \delta_{ij}, x_i, y_i)$, $1 \le i \le n$. The cured rate of individual i is

$$p_i^c = 1 - \sum_{j=1}^{J} p_{ij}, \qquad i = 1, 2, \ldots, n.$$

Let Δ_i be the contribution of individual i to the likelihood function based on the complete observations, we have

$$\Delta_i = \begin{cases} p_{ij} f_{ij}(t_i) & c_{ij} = \delta_{ij} = 1, \quad 1 \le j \le J, \\ p_i^c & c_i = \delta_i = 0, \\ p_{ij} S_{ij} & \delta_i = 0,\ c_{ij} = 1, \quad 1 \le j \le J. \end{cases}$$

The likelihood function and log-likelihood function on the basis of the complete observations are

$$\begin{aligned}
L_c &= \prod_{i=1}^{n} \Delta_i \\
&= \prod_{i=1}^{n} \left\{ \prod_{j=1}^{J} \big(p_{ij} f_{ij}(t_i)\big)^{c_{ij}\delta_{ij}} \big(1 - \sum_{j=1}^{J} p_{ij}\big)^{(1-c_i)(1-\delta_i)} \prod_{j=1}^{J} \big(p_{ij} S_{ij}(t_i)\big)^{c_{ij}(1-\delta_i)} \right\}
\end{aligned}$$

and

$$\begin{aligned}
l_c &= \sum_{i=1}^{n}\sum_{j=1}^{J} \left\{ c_{ij}\delta_{ij} \log\big(p_{ij} f_{ij}(t_i)\big) \right\} + \sum_{i=1}^{n} \left\{ (1-c_i)(1-\delta_i)\log(1 - \sum_{j=1}^{J} p_{ij}) \right\} \\
&\quad + \sum_{i=1}^{n}\sum_{j=1}^{J} \left\{ c_{ij}(1-\delta_i) \log\big(p_{ij} S_{ij}(t_i)\big) \right\}.
\end{aligned}$$

The log likelihood function l_c can be divided into two separate parts l_{c1} and l_{c2}, say, such that l_{c1} depends on p_{ij}'s only and l_{c2} depends on λ_{ij}'s only. So we can estimate the parameters corresponding to p_{ij}'s and λ_{ij}'s separately. The expressions of l_{c1} and l_{c2} are

$$\begin{aligned}
l_{c1} &= \sum_{i=1}^{n}\sum_{j=1}^{J} \left\{ c_{ij}\delta_{ij} \log p_{ij} \right\} + \sum_{i=1}^{n} \left\{ (1-c_i)(1-\delta_i)\log(1 - \sum_{j=1}^{J} p_{ij}) \right\} \\
&\quad + \sum_{i=1}^{n}\sum_{j=1}^{J} \left\{ c_{ij}(1-\delta_i) \log p_{ij} \right\} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{J} \left\{ c_{ij}(\delta_{ij} + 1 - \delta_i) \log p_{ij} \right\} + \sum_{i=1}^{n} \left\{ (1-c_i)(1-\delta_i)\log(1 - \sum_{j=1}^{J} p_{ij}) \right\}
\end{aligned}$$

and

$$\begin{aligned}
l_{c2} &= \sum_{i=1}^{n}\sum_{j=1}^{J} \left\{ c_{ij}\delta_{ij} \log\big(f_{ij}(t_i)\big) \right\} + \sum_{i=1}^{n}\sum_{j=1}^{J} \left\{ c_{ij}(1-\delta_i) \log\big(S_{ij}(t_i)\big) \right\} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{J} \left\{ c_{ij}\delta_{ij} \log\big(\lambda_{ij}(t_i)\big) \right\} + \sum_{i=1}^{n}\sum_{j=1}^{J} \left\{ c_{ij}(\delta_{ij} + 1 - \delta_i) \log\big(S_{ij}(t_i)\big) \right\}.
\end{aligned}$$

The expectations of l_{c1} and l_{c2} conditional on the incomplete observations and current estimates of the parameters are required to be calculated in the E-step in the EM algorithm. Let $g_{ij} = E[c_{ij}|(t_i, \delta_{ij}, x_i, y_i)]$, then

$$
\begin{aligned}
g_{ij} &= P\{c_{ij} = 1|\delta_{ij}, \delta_i\} \\
&= \delta_{ij} P\{c_{ij} = 1|\delta_{ij} = 1\} + (1-\delta_{ij})\delta_i P\{c_{ij} = 1|\delta_{ij} = 0,\ \delta_i = 1\} \\
&\quad + (1-\delta_i) P\{c_{ij} = 1|\delta_i = 0\} \\
&= \delta_{ij} + (1-\delta_i) \frac{P\{c_{ij} = 1, T_i^* > t_i\}}{P\{T_i^* > t_i\}} \\
&= \delta_{ij} + (1-\delta_i) \frac{p_{ij} S_{ij}(t_i)}{1 - \sum_{j=1}^J p_{ij} + \sum_{j=1}^J p_{ij} S_{ij}(t_i)}.
\end{aligned} \tag{4.3.9}
$$

So the conditional expectations of l_{c1} and l_{c2} are

$$
E[l_{c1}] = \sum_{i=1}^n \sum_{j=1}^J \{g_{ij} \log p_{ij}\} + \sum_{i=1}^n \left\{ (1-g_i) \log(1 - \sum_{j=1}^J p_{ij}) \right\}
$$

and

$$
E[l_{c2}] = \sum_{i=1}^n \sum_{j=1}^J \left\{ \delta_{ij} \log \left(h_{ij}(t_i) \right) \right\} + \sum_{i=1}^n \sum_{j=1}^J \left\{ g_{ij} \log \left(S_{ij}(t_i) \right) \right\},
$$

where $g_i = E[c_i] = \sum_{j=1}^J g_{ij}$. Then the EM Algorithm is ready to use. Let θ^q be the estimation of the parameter vector after the qth iteration. The iteration from θ^q to θ^{q+1} is comprised of the following two steps:

E-step: Calculate $E[l_{c1}(\theta^q)]$ and $E[l_{c2}(\theta^q)]$.

M-step: Find θ^{q+1} that maximizes $E[l_{c1}(\theta^q)]$ and $E[l_{c2}(\theta^q)]$.

The likelihood function based on the incomplete data is maximized by repeating the iterations from a chosen starting point.

Recall that to improve the approximation of the competing risks data with covariates and possible immunes, it is better to solve the following problems first:

(a) Are there immunes?

(b) Are the probabilities that individuals will eventually fail from a risk the same?

(c) Are the failure rates subject to a risk the same for all individuals?

(d) Are the failure rates of the risks of an individual the same with each other?

(e) Can we combine some intervals together? More specifically, is the exponential mixture model proper?

It is desired to investigate question (a) at the beginning of the data analysis. If there are immunes, the interior model should be adopted, otherwise the boundary model is preferred. Questions $(b) - (e)$ are studied for the simplicity of the model since the number of parameters decreases when the answers for the questions are "yes".

The test of the presence of immunes can only be carried out firstly. Once the existence of immunes is identified, the interior model or the boundary model can be chosen to fit the data accordingly. Thus for the model defined by (4.2.1), (4.2.2), (4.2.5) and (4.2.6), question (a) can be expressed by

H_{01}: $\gamma = 0$.

If H_{01} is accepted, the boundary model with $\gamma = 0$ is supposed to be proper. If H_{01} is rejected, the interior model may be more precise since the test implies the existence of immunes. For convenience of understanding, we describe the remaining tests in boundary model and in interior model separately. For the boundary case, questions $(b) - (e)$ can be expressed mathematically as:

H_{02}: $\beta_1 = \beta_2 = \cdots = \beta_{J-1} = 0$.

H_{03}: $\alpha_j = 0$, for $1 \leq j \leq J$.

H_{04}: $\varsigma_{1m} = \varsigma_{2m} = \cdots = \varsigma_{Jm}$ for $1 \leq m \leq M$ and $\alpha_1 = \alpha_2 = \cdots = \alpha_J$.

H_{05}: $\varsigma_{j1} = \varsigma_{j2} = \cdots = \varsigma_{jM}$ for $1 \leq j \leq J$.

Alternatively, for the model defined by (4.2.1), (4.2.2) and (4.2.4) questions $(b) - (e)$ can be formulated similarly as:

$H_{02}^{'}$: $\beta_1 = \beta_2 = \cdots = \beta_J = 0$.

$H_{03}^{'}$: $\alpha_j = 0$, for $1 \leq j \leq J$.

$H_{04}^{'}$: $\varsigma_{1m} = \varsigma_{2m} = \cdots = \varsigma_{Jm}$ for $1 \leq m \leq M$ and $\alpha_1 = \alpha_2 = \cdots = \alpha_J$.

$H_{05}^{'}$: $\varsigma_{j1} = \varsigma_{j2} = \cdots = \varsigma_{jM}$ for $1 \leq j \leq J$.

Let $d_n = -2\big(l_n(\hat{\theta}_0) - l_n(\hat{\theta}_a)\big)$ be the deviance statistic, where $\hat{\theta}_0$ and $\hat{\theta}_a$ are the estimates of the parameters under null and alternative hypotheses respectively. By Theorem 4.3.1, we have

$$d_n \xrightarrow{D} \begin{cases} N^2 I\{N \leq 0\} & \text{under} \quad H_{01}, \\ \chi^2_{(J-1)\pi_2} & \text{under} \quad H_{02}, \\ \chi^2_{J\pi_2} & \text{under} \quad H_{02}^{'}, \\ \chi^2_{J\pi_1} & \text{under} \quad H_{03} \quad \text{or} \quad H_{03}^{'}, \\ \chi^2_{(J-1)(M+\pi_1)} & \text{under} \quad H_{04} \quad \text{or} \quad H_{04}^{'}, \\ \chi^2_{J(M-1)} & \text{under} \quad H_{05} \quad \text{or} \quad H_{05}^{'}. \end{cases} \tag{4.3.10}$$

4.4 An Example of Real Life Data

We now illustrate the proposed approach with a subset of Bone Marrow Transplant Data listed in Tables 4.1 and 4.2, which have been studied by many authors, see for example Ibrahim, *et al* (2008) and Shu and Klein (2005). The common group (group 1) and AML low-risk group (group 2) are considered in our analysis. $N = 92$ patients who have received a transplant are exposed to two failure causes. δ_1 and δ_2 are the indicators of death with relapse and death without relapse, respectively. We can see from Table 4.3 that there are several differences between the two groups. The censoring of group 2 is heavier than that of group 1 and the proportion of failures caused by risk 2 among the deaths in group 2 is bigger that that in group 1. For the convenience of computation, the covariate is set to be $x = 0$ for group 1 and $x = 1$ for group 2.

The piecewise exponential mixture model specified in Section 4.2 is applied to the data. The cutpoints are set to be $\tau_1 = 800$ and $\tau_2 = 1600$ so that $M = 3$. We fit the data with several models based on different settings. Model 1 is the boundary model specified by (4.2.1), (4.2.2), (4.2.3), (4.2.5) and (4.2.6). The assumptions of model 2 are those of model 1 in addition with $\gamma = 0$. Model 3 is defined by (4.2.1), (4.2.2), (4.2.3) and (4.2.4). Model 4 is a specifical case of model 3 with $M = 1$. The estimates are developed by EM algorithm.

The models are assessed by the values of $-2 \log L$ and the performances are listed in Table 4.4. Two tests are carried out for the model selection and the details are in Table 4.5. According to the test of model 2 against model 1, the interior model should be adopted to approximate the data. The test of model 4 against model 3 suggests that an exponential mixture model is acceptable. The estimates developed by the exponential mixture model are presented in Table 4.6. The fitted model for this grouped data displayed in Table 4.7. It is easy to see that the cured rate of group 2 is higher than that of group 1. This is reasonable since group 2 is AML low-risk.

Figure 4.1 and Figure 4.2 show the survival curves developed by the exponential mixture model, the approach of Kaplan and Meier (1958) and the method of Larson and Dinse (1985). The curves developed by our approach are similar to those by the method of Kaplan and Meier (1958) for both groups. The approach of Larson and Dinse (1985) tends to over-estimate the survival up to about 1100 days for group 1 and 1400 days for group 2, and then predict shorter survival than the other two methods. The main reason is that the existences of immunes is ignored in the approach of Larson and Dinse (1985).

Compared with the previous models, our approach in the analysis of the BMT data has several advantages:

1. It is more flexible to fit the data and model the real-life situation by relaxing the restriction that all individuals under the study must eventually fail from one of the risk under consideration.

2. It is able to test the presence of long-term survivors and thus select the appropriate model – interior or boundary.

3. It provides more accurate estimates, especially if there is a substantial proportion of long-term survivors.

We also carried out the boundary hypothesis testing on a subset of Standford Heart Transplant Data (Crowley and Hu, 1997). This set of data has been studied by Larson and Dinse (1985) with the boundary model. The test shows that the existence of immunes is insignificant (deviance = 0.24). Hence the work of Larson and Dince (1985) is supported by our results.

Table 4.1: Bone Marrow Transplant Data for Group 1

t	δ_1	δ_2	t	δ_1	δ_2	t	δ_1	δ_2
2081	0	0	1167	0	0	1279	1	0
1602	0	0	418	0	1	110	1	0
1496	0	0	417	1	0	243	1	0
1462	0	0	276	0	1	86	0	1
1433	0	0	156	1	0	466	0	1
1377	0	0	781	1	0	262	1	0
1330	0	0	172	0	1	162	1	0
996	0	0	487	0	1	262	1	0
226	0	0	716	1	0	1	0	1
1199	0	0	194	0	1	107	0	1
1111	0	0	371	1	0	269	1	0
530	0	0	526	0	1	350	0	0
1182	0	0	122	0	1			

Table 4.2: Bone Marrow Transplant Data for Group 2

t	δ_1	δ_2	t	δ_1	δ_2	t	δ_1	δ_2
2569	0	0	1527	0	0	288	0	1
2506	0	0	1324	0	0	522	1	0
2409	0	0	957	0	0	79	0	1
2218	0	0	932	0	0	1156	1	0
1857	0	0	847	0	0	583	1	0
1829	0	0	848	0	0	48	0	1
1562	0	0	1850	0	0	431	1	0
1470	0	0	1843	0	0	1074	0	1
1363	0	0	1535	0	0	393	1	0
1030	0	0	1447	0	0	10	0	1
860	0	0	1384	0	0	53	0	1
1258	0	0	414	0	1	80	0	1
2246	0	0	2204	0	1	35	0	1
1870	0	0	1063	0	1	1499	0	0
1799	0	0	481	0	1	704	0	1
1709	0	0	105	0	1	653	1	0
1674	0	0	641	0	1	222	1	0
1568	0	0	390	0	1	1356	0	0

Table 4.3: Data Analysis

Group	Total No.	Risk 1	Risk 2	Censoring
1	38	12 (31.58%)	11 (28.95%)	15 (39.47%)
2	54	7 (12.96%)	16 (29.63%)	31 (54.71%)
Total	92	19 (20.65%)	27 (29.35%)	46 (50.00%)

Table 4.4: Performances of Different Models

Model	Type	M	Cutpoints	-2Log-L
1	boundary	3	800, 1600	816.4664
2	boundary, $\gamma = 0$	3	800, 1600	821.6395
3	interior	3	800, 1600	816.3791
4	interior	1	NA	821.6194

Table 4.5: Results of The Tests

H_0	H_1	Deviance	Distribution	Critical value	Result
model 2	model 1	5.1731	$N^2 I\{N \leq 0\}$	2.71(95%)	reject
model 4	model 3	5.2403	χ^2_4	7.78(90%)	accept

Table 4.6: Estimation of The Parameters

Parameter	α_1	α_2	ς_1	ς_2	ξ_1	ξ_2	β_1	β_2
Estimate	-0.3974	-0.7728	-6.3739	-5.6511	0.1589	-0.0544	-1.3367	-0.3969

Table 4.7: Fitted Model

Group i	p_{i1}	p_{i2}	p_i^c	λ_{i1}	λ_{i2}
1	0.3758	0.3036	0.3206	0.0017	0.0035
2	0.1978	0.2626	0.5396	0.0011	0.0016

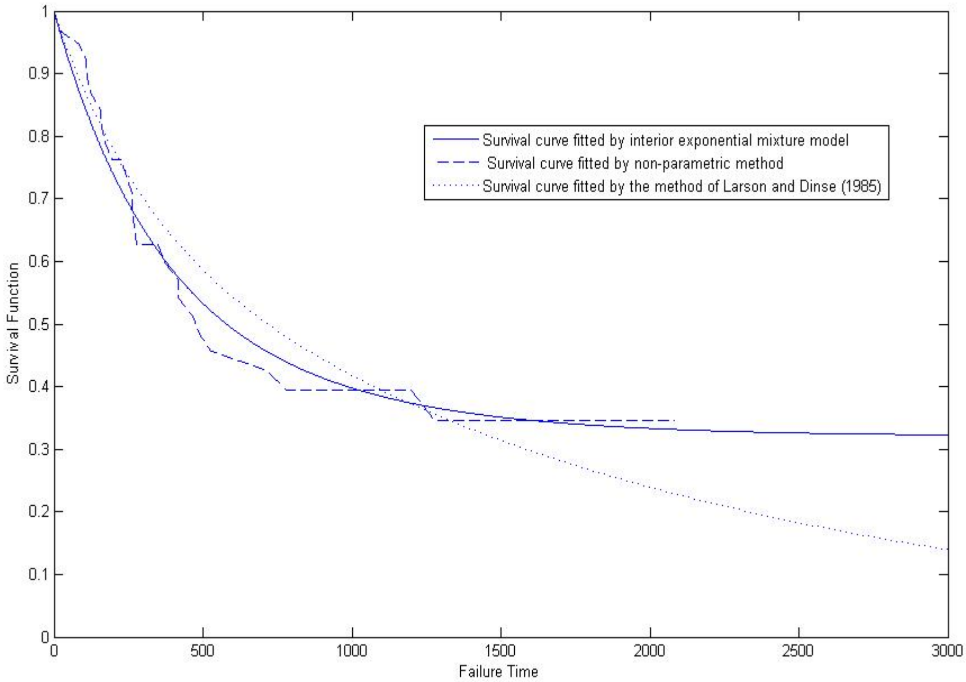

Figure 4.1: Comparison of Survival Curves for Group 1

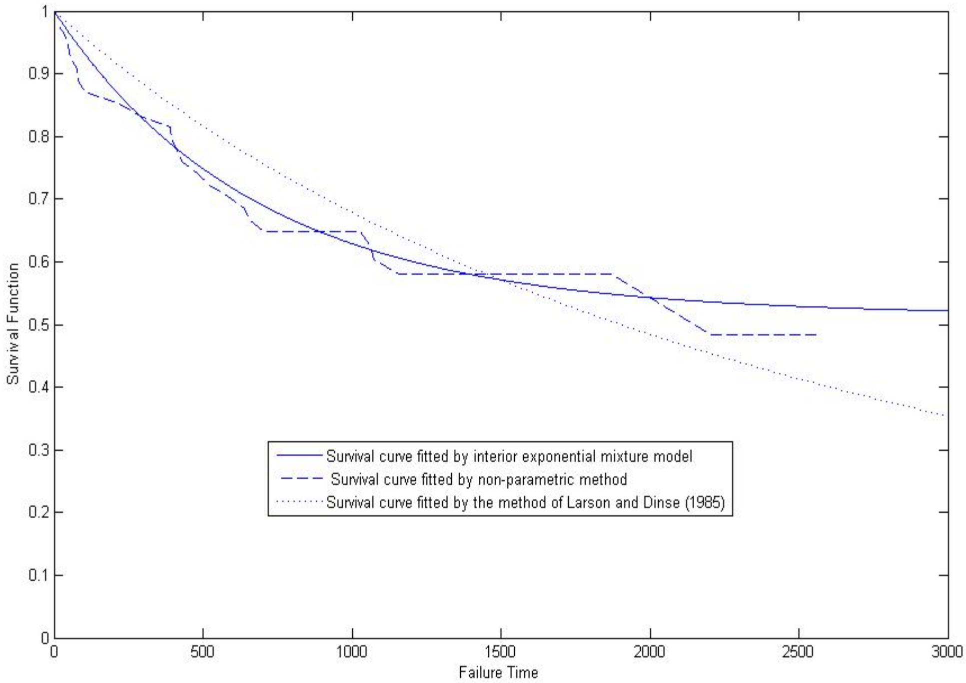

Figure 4.2: Comparison of Survival Curves for Group 2

4.5 Concluding Remarks

In this chapter, we discussed the application of a piecewise exponential mixture model and EM algorithm on competing risks data with covariates and possible immunes. Two real life data sets are investigated and the results favorite our approach.

The theoretical results of the general mixture model provide the basic large-sample foundations for competing risks data analysis in the possible presence of immunes. But it is difficult to identify the distribution of the failure time conditional on a give risk at the origin of a study. The piecewise exponential mixture model provide a convenient process to specify the hazard function of the failure time conditional on a give risk by selecting the cutpoints appropriately. In this way, the theoretical results developed in Chapter 3 could be applied easily in practical survival analysis.

4.6 Proofs

As the proofs are similar for the interior case and boundary case of the model, we illustrate the details of the proof based on the boundary model, in which

$$\theta = (\varsigma_1^T, \ldots, \varsigma_J^T, \alpha_1^T, \ldots, \alpha_J^T, \beta_1^T, \ldots, \beta_{J-1}^T, \xi_1^T, \ldots, \xi_{J-1}^T, \gamma)^T$$

and

$$H_i = \begin{bmatrix} I_{JM} & 0 & 0 & 0 & 0 \\ 0 & I_J \otimes x_i & 0 & 0 & 0 \\ 0 & 0 & I_{J-1} \otimes y_i & 0 & 0 \\ 0 & 0 & 0 & I_{J-1} & 0 \\ 0 & 0 & 0 & 0 & I_1 \end{bmatrix}. \tag{4.6.11}$$

Lemma 4.6.1 *In the piecewise exponential mixture model specified in Section 4.2, if the censorship and covariates are bounded and condition (E) holds, $Var\{f_i^{lk}(\theta_0)\}$ are uniformly bounded for $1 \le i \le n$ and $1 \le l,\, k \le J(M+3)-1$.*

Proof. First we derive some useful equations. For $1 \le i \le n$, $1 \le j, r \le J$, $1 \le l, w \le J-1$ and $1 \le m, k \le M$,

$$s_i^{M(j-1)+m} = 1_{\{t_i \in \mathcal{I}_m\}} \delta_{ij}(1 - t_i \lambda_{ij}^m) - \frac{(1-\delta_i) p_{ij} \lambda_{ij}^m}{1 - F_i(t_i)} \frac{\partial F_{ij}(t_i)}{\partial \lambda_{ij}^m},$$

$$s_i^{JM+j} = \delta_{ij}\big(1 - t_i \lambda_{ij}(t_i)\big) - \sum_{m=1}^{M} \frac{(1-\delta_i) p_{ij} \lambda_{ij}^m}{1 - F_i(t_i)} \frac{\partial F_{ij}(t_i)}{\partial \lambda_{ij}^m},$$

$$s_i^{J(M+1)+l} = s_i^{J(M+2)-1+l} = \delta_{il} - p_{il} + \frac{(1-\delta_i) p_{il} (1 - F_{il}(t_i))}{1 - F_i(t_i)},$$

$$s_i^{J(M+3)-1} = -p_{iJ} + \frac{(1-\delta_i) p_{iJ}}{1 - F_i(t_i)},$$

$$\begin{aligned} f_i^{M(j-1)+m, M(r-1)+k} =& 1_{\{t_i \in \mathcal{I}_m, j=r, m=k\}} \delta_{ij} t_i \lambda_{ij}^m + 1_{\{j=r, m=k\}} \frac{(1-\delta_i) p_{ij} \lambda_{ij}^m}{1 - F_i(t_i)} \frac{\partial F_{ij}(t_i)}{\partial \lambda_{ij}^m} \\ &+ 1_{\{j=r, m=k\}} \frac{(1-\delta_i) p_{ij} (\lambda_{ij}^m)^2}{1 - F_i(t_i)} \frac{\partial^2 F_{ij}(t_i)}{(\partial \lambda_{ij}^m)^2} \\ &+ \frac{(1-\delta_i) p_{ij} p_{ir} \lambda_{ij}^m \lambda_{ir}^k}{\big(1 - F_i(t_i)\big)^2} \frac{\partial F_{ij}(t_i)}{\partial \lambda_{ij}^m} \frac{\partial F_{ir}(t_i)}{\partial \lambda_{ir}^k}, \end{aligned}$$

$$\begin{aligned}
f_i^{M(j-1)+m,JM+r} =& 1_{\{t_i\in\mathcal{I}_m\}}\delta_{ij}t_i\lambda_{ij}^m + \frac{(1-\delta_i)p_{ij}\lambda_{ij}^m}{1-F_i(t_i)}\frac{\partial F_{ij}(t_i)}{\partial\lambda_{ij}^m} + \frac{(1-\delta_i)p_{ij}(\lambda_{ij}^m)^2}{1-F_i(t_i)}\frac{\partial^2 F_{ij}(t_i)}{(\partial\lambda_{ij}^m)^2} \\
&+ \frac{(1-\delta_i)p_{ij}p_{ir}\lambda_{ij}^m}{\big(1-F_i(t_i)\big)^2}\frac{\partial F_{ij}(t_i)}{\partial\lambda_{ij}^m}\frac{\partial F_{ir}(t_i)}{\partial\alpha_r},
\end{aligned}$$

$$\begin{aligned}
f_i^{M(j-1)+m,J(M+1)+l} =& f_i^{M(j-1)+m,J(M+2)-1+l} \\
=& \frac{(1-\delta_i)p_{ij}\lambda_{ij}^m}{1-F_i(t_i)}\frac{\partial F_{ij}(t_i)}{\partial\lambda_{ij}^m}\left\{1_{\{l=j\}} - \frac{p_{il}\big(1-F_{il}(t_i)\big)}{1-F_i(t_i)}\right\},
\end{aligned}$$

$$f_i^{M(j-1)+m,J(M+3)-1} = -\frac{(1-\delta_i)\lambda_{ij}^m p_{ij}p_{iJ}}{\big(1-F_i(t_i)\big)^2}\frac{\partial F_{ij}(t_i)}{\partial\lambda_{ij}^m},$$

$$f_i^{JM+j,JM+r} = 1_{\{r=j\}}\delta_{ij}t_i\lambda_{ij}(t_i) + 1_{\{r=j\}}\frac{(1-\delta_i)p_{ij}}{1-F_i(t_i)}\frac{\partial^2 F_{ij}(t_i)}{(\partial\alpha_j)^2} + \frac{(1-\delta_i)p_{ij}p_{ir}}{\big(1-F_i(t_i)\big)^2}\frac{\partial F_{ij}(t_i)}{\partial\alpha_j}\frac{\partial F_{ir}(t_i)}{\partial\alpha_r},$$

$$\begin{aligned}
f_i^{JM+j,J(M+1)+l} =& f_i^{JM+j,J(M+2)-1+l} \\
=& 1_{\{l=j\}}\frac{(1-\delta_i)p_{il}}{1-F_i(t_i)}\frac{\partial F_{ij}(t_i)}{\partial\alpha_j} - \frac{(1-\delta_i)\big(1-F_{il}(t_i)\big)p_{il}p_{ij}}{\big(1-F_i(t_i)\big)^2}\frac{\partial F_{ij}(t_i)}{\partial\alpha_j},
\end{aligned}$$

$$f_i^{JM+j,J(M+3)-1} = -\frac{(1-\delta_i)p_{ij}p_{iJ}}{\big(1-F_i(t_i)\big)^2}\frac{\partial F_{ij}(t_i)}{\partial\alpha_j},$$

$$\begin{aligned}
f_i^{J(M+1)+l,J(M+1)+w} =& f_i^{J(M+1)+l,J(M+2)-1+w} \\
=& f_i^{J(M+2)-1+l,J(M+1)+w} = f_i^{J(M+2)-1+l,J(M+2)-1+w} \\
=& 1_{\{w=l\}}p_{il} - p_{il}p_{iw} - 1_{\{l=w\}}\frac{(1-\delta_i)\big(1-F_{il}(t_i)\big)p_{il}}{1-F_i(t_i)} \\
&+ \frac{(1-\delta_i)\big(1-F_{il}(t_i)\big(1-F_{iw}(t_i)\big)p_{il}p_{iw}}{\big(1-F_i(t_i)\big)^2},
\end{aligned}$$

$$f_i^{J(M+1)+l,J(M+3)-1} = f_i^{J(M+2)-1+l,J(M+3)-1} = -p_{il}p_{iJ} + \frac{(1-\delta_i)\big(1-F_{il}(t_i)\big)p_{il}p_{iJ}}{\big(1-F_i(t_i)\big)^2},$$

$$f_i^{J(M+3)-l,J(M+3)-1} = -(p_{iJ})^2 + \frac{(1-\delta_i)(p_{iJ})^2}{\big(1-F_i(t_i)\big)^2},$$

where $1_{\{\cdot\}}$ is an indicator function and

$$\frac{\partial F_{ij}(t_i)}{\partial \lambda_{ij}^m} = 1_{\{t_i > \tau_{m-1}\}} \left\{ (t_i \wedge \tau_m) e^{-\lambda_{ij}^m (t_i \wedge \tau_m)} - \tau_{m-1} e^{-\lambda_{ij}^m \tau_{m-1}} \right\},$$

$$\frac{\partial^2 F_{ij}(t_i)}{(\partial \lambda_{ij}^m)^2} = 1_{\{t_i > \tau_{m-1}\}} \left\{ -\big((t_i \wedge \tau_m)\big)^2 e^{-\lambda_{ij}^m (t_i \wedge \tau_m)} + \tau_{m-1}^2 e^{-\lambda_{ij}^m \tau_{m-1}} \right\},$$

$$\frac{\partial F_{ij}(t_i)}{\partial \alpha_j} = \sum_{m=1}^{M} 1_{\{t_i > \tau_{m-1}\}} \lambda_{ij}^m \left\{ (t_i \wedge \tau_m) e^{-\lambda_{ij}^m (t_i \wedge \tau_m)} - \tau_{m-1} e^{-\lambda_{ij}^m \tau_{m-1}} \right\},$$

$$\begin{aligned}\frac{\partial^2 F_{ij}(t_i)}{(\partial \alpha_j)^2} &= \sum_{m=1}^{M} 1_{\{t_i > \tau_{m-1}\}} \lambda_{ij}^m \left\{ (t_i \wedge \tau_m) e^{-\lambda_{ij}^m (t_i \wedge \tau_m)} - \tau_{m-1} e^{-\lambda_{ij}^m \tau_{m-1}} \right\} \\ &\quad + 1_{\{t_i > \tau_{m-1}\}} (\lambda_{ij}^m)^2 \left\{ -\big((t_i \wedge \tau_m)\big)^2 e^{-\lambda_{ij}^m (t_i \wedge \tau_m)} + \tau_{m-1}^2 e^{-\lambda_{ij}^m \tau_{m-1}} \right\}.\end{aligned}$$

Let u_i be the censoring time associated with individual i,

$$\begin{aligned} 1 - F_{i0}(u_i) &\geq \min_{1\leq j\leq J,\, 1\leq m\leq M} \{e^{-\lambda_{ij0}^m u_i}\} \\ &\geq \min_{1\leq j\leq J,\, 1\leq m\leq M} \{e^{-\lambda_{ij0}^m u_i}\} \\ &\geq \exp\{-\sup_{1\leq j\leq J,\, 1\leq m\leq M} \lambda_{ij0}^m \sup_{1\leq i\leq n} u_i\} \\ &= K^*, \end{aligned} \tag{4.6.12}$$

where K^* is a positive constant.

Thus for $1 \leq l,\, k \leq J(M+3)-1$, $Var\{f_i^{lk}(\theta_0)\} \leq E[\big(f_i^{lk}(\theta_0)\big)^2]$ and $\big(f_i^{lk}(\theta_0)\big)^2$ are linear combinations of the following terms with finite coefficients:

$$1_{\{t_i \in \mathcal{I}_m\}} \delta_{ij} \left\{ t_i \lambda_{ij0}^m \right\}^2,$$

$$\left\{\lambda_{ij0}^m (t_i \wedge \tau_m) e^{-\lambda_{ij0}^m (t_i\wedge\tau_m)}\right\}^2, \qquad \left\{\lambda_{ij0}^m \tau_{m-1} e^{-\lambda_{ij0}^m \tau_{m-1}}\right\}^2,$$

$$(\lambda_{ij0}^m \lambda_{ir0}^k)^2 (t_i \wedge \tau_m)\tau_{k-1} e^{-\lambda_{ij0}^m (t_i\wedge\tau_m)} e^{-\lambda_{ir0}^k \tau_{k-1}},$$

$$(\lambda_{ij0}^m)^2 (\lambda_{ir0}^k)^2 \left\{(t_i \wedge \tau_m)(t_i \wedge \tau_k) e^{-\lambda_{ij0}^m (t_i\wedge\tau_m)} e^{-\lambda_{ir0}^k (t_i\wedge\tau_k)}\right\}^2,$$

$$(\lambda_{ij0}^m)^2 (\lambda_{ir0}^k)^2 \left\{\tau_{m-1}\tau_{k-1} e^{-\lambda_{ij0}^m \tau_{m-1}} e^{-\lambda_{ir0}^k \tau_{k-1}}\right\}^2,$$

$$(\lambda_{ij0}^m)^2 (\lambda_{ir0}^k)^2 (t_i \wedge \tau_m)^2 (t_i \wedge \tau_k)\tau_{k-1} e^{-2\lambda_{ij0}^m (t_i\wedge\tau_m)} e^{-\lambda_{ir0}^k (t_i\wedge\tau_k)} e^{-\lambda_{ir0}^k \tau_{k-1}},$$

$$(\lambda_{ij0}^m)^2 (\lambda_{ir0}^k)^2 (t_i \wedge \tau_m)(t_i \wedge \tau_k)\tau_{m-1}\tau_{k-1} e^{-\lambda_{ij0}^m (t_i\wedge\tau_m)} e^{-\lambda_{ir0}^k (t_i\wedge\tau_k)} e^{-\lambda_{ij0}^m \tau_{m-1}} e^{-\lambda_{ir0}^k \tau_{k-1}},$$

$$(\lambda_{ij0}^m)^2 (\lambda_{ir0}^k)^2 (t_i \wedge \tau_m)\tau_{m-1}\tau_{k-1}^2 e^{-\lambda_{ij0}^m (t_i\wedge\tau_m)} e^{-\lambda_{ij0}^m \tau_{m-1}} e^{-2\lambda_{ir0}^k \tau_{k-1}}.$$

Noting that for $1 \leq i \leq n$, $1 \leq j \leq J$ and $1 \leq m \leq M$, we have

$$\begin{aligned} E[1_{\{t_i \in \mathcal{I}_m\}} \delta_{ij} (t_i \lambda_{ij0}^m)^2] &= p_{ij0} E\left\{\int_{\tau_{m-1}}^{u_i} (\lambda_{ij0}^m t)^2 \lambda_{ij0}^m e^{-\lambda_{ij0}^m t} dt\right\} \\ &\leq \sup\left\{\int_0^\infty (\lambda_{ij0}^m t)^2 \lambda_{ij0}^m e^{-\lambda_{ij0}^m t} dt\right\} \\ &= \sup\left\{\int_0^\infty y^2 e^{-y} dy\right\} \\ &= 2 \end{aligned} \tag{4.6.13}$$

and for any positive integer $a < \infty$ we have

$$\sup_{x>0} x^a e^{-x} = a^a e^{-a} < \infty, \tag{4.6.14}$$

we can argue that the above terms that constitute $\left(f_i^{lk}(\theta_0)\right)^2$ are uniformly bounded. It follows that $Var\{f_i^{lk}(\theta_0)\}$ are uniformly bounded. ■

Lemma 4.6.2 *In the piecewise exponential mixture model specified in Section 4.2 with bounded censorship and covariates, if condition (B1) and condition (E) hold, condition (B3) holds.*

Proof. For $1 \le l,\, k \le J(M+3)-1$ and $1 \le i \le n$, let

$$q_{ilk}^{M(j-1)+m} = \lambda_{ij}^{m} \frac{\partial f_i^{lk}(\theta)}{\partial \lambda_{ij}^{m}}, \qquad 1 \le j \le J,\, 1 \le m \le M,$$

$$q_{ilk}^{JM+j} = \sum_{m=1}^{M} \lambda_{ij}^{m} \frac{\partial f_i^{lk}(\theta)}{\partial \lambda_{ij}^{m}}, \qquad 1 \le j \le J,$$

$$q_{ilk}^{J(M+1)+w} = q_{ilk}^{J(M+2)-1+w} = p_{iw}\left\{ \frac{\partial f_i^{lk}(\theta)}{\partial p_{iw}} - \sum_{j=1}^{J} p_{ij} \frac{\partial f_i^{lk}(\theta)}{\partial p_{ij}} \right\}, \qquad 1 \le w \le J-1,$$

$$q_{ilk}^{J(M+3)-1} = -p_{iJ} \sum_{j=1}^{J} p_{ij} \frac{\partial f_i^{lk}(\theta)}{\partial p_{ij}}.$$

We will then show that there exists a constant $\mathcal{K} > 0$ such that

$$E\{\sup_{\theta \in \mathcal{N}_n(A)} |q_{ilk}(\theta)|\} \le \mathcal{K} \tag{4.6.15}$$

for $\theta \in \mathcal{N}_n(A)$, $1 \le l,\, k \le J(M+3)-1$ and $1 \le i \le n$. For simplicity, we provide the proof for $l = k = 1$. It suffices to show that there exist $\mathcal{K}_1 > 0$ and $\mathcal{K}_2 > 0$ such that

$$E\{\sup_{\theta \in \mathcal{N}_n(A)} 1_{\{t_i \in \mathcal{I}_1\}} \delta_{i1} t_i \lambda_{i1}^{1}\} \le \mathcal{K}_1 \tag{4.6.16}$$

and for $0 \le a_1,\, a_2,\, a_3,\, a_4 \le 3$,

$$E\left\{ \sup_{\theta \in \mathcal{N}_n(A)} (1-\delta_i) \frac{(\lambda_{i1}^{1} t_i)^{a_1} e^{-\lambda_{i1}^{1} t_i} (\lambda_{i1}^{1} \tau_m)^{a_2} e^{-\lambda_{i1}^{1} \tau_m} (\lambda_{i1}^{1} \tau_{m-1})^{a_3} e^{-\lambda_{i1}^{1} \tau_{m-1}}}{\big(1 - F_i(t_i)\big)^{a_4}} \right\} \le \mathcal{K}_2. \tag{4.6.17}$$

Noting that for $\theta \in \mathcal{N}_n(A)$,

$$
\begin{aligned}
|H_i^T(\theta-\theta_0)|^2 &= |H_i^T \mathcal{Q}_n^{-1/2}\mathcal{Q}_n^{1/2}(\theta-\theta_0)|^2 \\
&\le |\mathcal{Q}_n^{1/2}(\theta-\theta_0)\,|^2\;|\lambda_{\max}\{\mathcal{Q}_n^{-1/2}H_iH_i^T\mathcal{Q}_n^{1/2}\} \\
&\le A^2 tr\{H_i^T\mathcal{Q}_n^{-1}H_i\}, \qquad (4.6.18)
\end{aligned}
$$

by condition (B1) and condition (E), we have, for $\theta \in \mathcal{N}_n(A)$ and $0 \le v < \infty$,

$$
\begin{aligned}
|(\lambda_{ij}^m)^v-(\lambda_{ij0}^m)^v| &= (\lambda_{ij0}^m)^v|v\big(1-e^{\varsigma_j^m-\varsigma_{j0}^m+x_i^T(\alpha_j-\alpha_{j0})}\big)| \\
&= (\lambda_{ij0}^m)^v|v\big(\varsigma_j^m-\varsigma_{j0}^m+x_i^T(\alpha_j-\alpha_{j0})\big)|\sum_{r=1}^{\infty}\frac{\big\{v\big(\varsigma_j^m-\varsigma_{j0}^m+x_i^T(\alpha_j-\alpha_{j0})\big)\big\}^{r-1}}{r!} \\
&\le (\lambda_{ij0}^m)^v|v\big(\varsigma_j^m-\varsigma_{j0}^m+x_i^T(\alpha_j-\alpha_{j0})\big)|\sum_{r=0}^{\infty}\frac{|v\big(\varsigma_j^m-\varsigma_{j0}^m+x_i^T(\alpha_j-\alpha_{j0})\big)|^r}{(r)!} \\
&= (\lambda_{ij0}^m)^v|v\big(\varsigma_j^m-\varsigma_{j0}^m+x_i^T(\alpha_j-\alpha_{j0})\big)|\exp\{v|\varsigma_j^m-\varsigma_{j0}^m+x_i^T(\alpha_j-\alpha_{j0})|\} \\
&\le (\lambda_{ij0}^m)^v vA\left\{tr\{H_i^T\mathcal{Q}_n^{-1}H_i\}\right\}^{1/2}\exp\left\{vA\left\{tr\{H_i^T\mathcal{Q}_n^{-1}H_i\}\right\}^{1/2}\right\} \\
&\longrightarrow 0 \qquad (n\longrightarrow\infty). \qquad (4.6.19)
\end{aligned}
$$

Similarly, for $\theta \in \mathcal{N}_n(A)$, we can drive

$$
|p_{ij}-p_{ij0}| \longrightarrow 0 \qquad (n\longrightarrow\infty). \qquad (4.6.20)
$$

Hence, for n large enough,

$$
|\lambda_{ij}^m-\lambda_{ij0}^m| \le \inf_{1\le r\le J,\,1\le k\le M}\lambda_{ir0}^k \le \lambda_{ij0}. \qquad (4.6.21)
$$

It follows that

$$
\begin{aligned}
E\{\sup_{\theta\in\mathcal{N}_n(A)} 1_{\{t_i\in\mathcal{I}_1\}}\delta_{i1}t_i\lambda_{i1}^1\} &\le 2E[1_{\{t_i\in\mathcal{I}_1\}}\delta_{i1}t_i\lambda_{i10}^1] \\
&\le 2p_{i10}E\left\{\int_0^{u_i}\lambda_{i10}^1 t\lambda_{i10}^1 e^{-\lambda_{i10}^1 t}dt\right\} \\
&\le 2\sup\left\{\int_0^{\infty}(\lambda_{i10}^1 t)^2\lambda_{i10}^1 e^{-\lambda_{i10}^1 t}dt\right\} \\
&= 2\sup\left\{\int_0^{\infty} ye^{-y}dy\right\} = 2
\end{aligned}
\tag{4.6.22}
$$

and so (4.6.16) holds and it leaves (4.6.17) to be verified. By (4.6.14), there exists $\mathcal{K}_3 > 0$ such that

$$
(\lambda_{ij}^m)^b\left\{\frac{\partial F_{ij}(t_i)}{\partial \lambda_{ij}^m}\right\} \le \mathcal{K}_3. \tag{4.6.23}
$$

Also for $\theta \in \mathcal{N}_n(A)$,

$$
\begin{aligned}
&|F_i(u_i) - F_{i0}(u_i)| \\
&\le \sum_{j=1}^{J}|p_{ij}\big(F_{ij}(u_i) - F_{ij0}(u_i)\big)| + \sum_{j=1}^{J}|F_{ij0}(u_i)(p_{ij} - p_{ij0})| \\
&\le \sum_{j=1}^{J}\sum_{m=1}^{M}1_{\{u_i>\tau_{m-1}\}}|e^{-\lambda_{ij}^m\tau_{m-1}} - e^{-\lambda_{ij0}^m\tau_{m-1}}| + |e^{-\lambda_{ij}^m(\tau_m\wedge u_i)} - \sum_{j=1}^{J}\sum_{m=1}^{M}e^{-\lambda_{ij0}^m(\tau_m\wedge u_i)}| \\
&\quad + \sum_{j=1}^{J}|F_{ij0}(u_i)(p_{ij} - p_{ij0})| \longrightarrow 0 \qquad (n\longrightarrow\infty)
\end{aligned}
$$

by (4.6.19) and(4.6.20). Thus for $\theta \in \mathcal{N}_n(A)$,

$$
|\{1 - F_i(u_i)\} - \{1 - F_{i0}(u_i)\}| \longrightarrow 0 \qquad (n\longrightarrow\infty). \tag{4.6.24}
$$

$$
\begin{aligned}
&E\left\{\sup_{\theta\in\mathcal{N}_n(A)}(1-\delta_i)\frac{(\lambda_{i1}^1 t_i)^{a_1}e^{-\lambda_{i1}^1 t_i}(\lambda_{i1}^1\tau_m)^{a_2}e^{-\lambda_{i1}^1\tau_m}(\lambda_{i1}^1\tau_{m-1})^{a_3}e^{-\lambda_{i1}^1 t_{m-1}}}{\big(1 - F_i(t_i)\big)^{a_4}}\right\} \\
&= E\left\{\sup_{\theta\in\mathcal{N}_n(A)}\frac{(\lambda_{i1}^1 u_i)^{a_1}e^{-\lambda_{i1}^1 u_i}(\lambda_{i1}^1\tau_m)^{a_2}e^{-\lambda_{i1}^1\tau_m}(\lambda_{i1}^1\tau_{m-1})^{a_3}e^{-\lambda_{i1}^1\tau_{m-1}}}{\big(1 - F_i(u_i)\big)^{a_4-1}}\right\}.
\end{aligned}
\tag{4.6.25}
$$

So (4.6.12), (4.6.23), (4.6.24) and (4.6.25) imply that (4.6.17) holds and further imply that (4.6.15) holds.

Hence for $1 \le l,\, k \le J(M+3)-1$, by Taylor expansion, (4.6.18) and (4.6.15),

$$\begin{aligned} E\left\{\sup_{\theta\in\mathscr{N}_n(A)} |f_i^{lk}(\theta) - f_i^{lk}(\theta_0)|\right\} &= E\left\{\sup_{\theta\in\mathscr{N}_n(A)} |(\theta-\theta_0)^T \frac{\partial f_i^{li}}{\partial\theta}(\theta^*)|\right\} \\ &= E\left\{\sup_{\theta\in\mathscr{N}_n(A)} |(\theta-\theta_0)^T H_i q_{ilk}(\theta^*)|\right\} \\ &\le A\{tr(H_I^T \mathcal{Q}_n^{-1} H_i)\}^{1/2} E\left\{\sup_{\theta\in\mathscr{N}_n(A)} |q_{ilk}(\theta)|\right\} \\ &\le \mathcal{K} A\{tr(H_I^T \mathcal{Q}_n^{-1} H_i)\}^{1/2}, \end{aligned}$$

which implies that condition (B3) holds.

Proof of Theorem 4.3.1. By Lemma 4.6.1 and 4.6.2, condition (B1) together with condition (E) imply that condition (B2) and condition (B3) are satisfied. It suffices to verify condition (B4) to prove Theorem 4.3.1. Let

$$\mathscr{X}_{jm}(t_i) = \big(\lambda_{ij0}^m (t_i \wedge \tau_m)\big)^{b_1} e^{-a_1 \lambda_{ij0}^m (t_i\wedge\tau_m)} \big(\lambda_{ij0}^m \tau_{m-1}\big)^{b_2} e^{-a_2\lambda_{ij0}^m \tau_{m-1}},$$

where $0 \le a_1,\, a_2,\, b_1,\, b_2 \le 4$. For $1 \le r \le J(M+3)-1$, $\{s_i^r(\theta_0)\}^4$ are linear combinations of the following terms:

$$1_{\{t_i\in\mathcal{I}_m\}}\delta_{ij0}(1 - t_i\lambda_{ij0}^m)^4, \qquad \frac{(1-\delta_i)}{\big(1-F_i(t_i)\big)^{b_3}} \mathscr{X}_{jm}(t_i)\mathscr{X}_{jk}(t_i), \qquad 1 \le b_3 \le 4, \quad (4.6.26)$$

for $1 \leq j \leq J$ and $1 \leq m \leq M$. We only need to prove that the expectations of the items of (4.6.26) are uniformly bounded.

$$\begin{aligned}
E\{1_{\{t_i \in \mathcal{I}_m\}}\delta_{ij0}(1 - t\lambda_{ij0}^m)^4\} &= p_{ij0}E\left\{\int_{\tau_{m-1}}^{u_i}(1 - t\lambda_{ij0}^m)^4\lambda_{ij0}^m e^{-\lambda_{ij0}^m t}dt\right\} \\
&\leq \sup\left\{\int_0^\infty (1 - t\lambda_{ij0}^m)^4\lambda_{ij0}^m e^{-\lambda_{ij0}^m t}dt\right\} \\
&= \sup\left\{\int_0^\infty (1 - y)^4 e^{-y}dy\right\} \\
&= 9,
\end{aligned}$$

which implies that $E\{1_{\{t_i \in \mathcal{I}_m\}}\delta_{ij0}(1 - t\lambda_{ij0}^m)^4\}$ is uniformly bounded.

$$E\left\{\frac{(1-\delta_i)}{\big(1 - F_i(t_i)\big)^{b_3}}\mathscr{X}_{jm}(t_i)\mathscr{X}_{jk}(t_i)\right\} = E\left\{\frac{1}{\big(1 - F_i(u_i)\big)^{b_3 - 1}}\mathscr{X}_{jm}(u_i)\mathscr{X}_{jk}(u_i)\right\}$$

is uniformly bounded by (4.6.12) and (4.6.14). Hence condition (B4) holds and Theorem 4.3.1 is proved. ■

Chapter 5

A Proportional Hazards Model

5.1 Introduction

Methodologies based on the hazard functions (Cox, 1972) have been popular for decades because they can approximate survival data more naturally than multivariate survival functions do. Among them Cox's proportional hazards approach has been the most widely used. On the other hand, the methodology on the basis of martingale counting processes has been developed in the modern literature, for it can provide a unified and efficient way to give proofs. See, for example, Aalen (1975), Fleming and Harrington (1991), Andersen and Gill (1982), Andersen, Borgan, Gill and Keiding (1993) and Zhao and Zhou (2006).

Our work is motivated by the fact that studies in proportional hazards models for survival data that combine competing risks, covariates and immunes have not been reported. To narrow this gap, a semi-parametric approach is proposed and modern techniques of counting processes and martingales are adopted to derive the existence, consistency and asymptotic normality of the estimators. The advantage of our approach over the parametric model is that we can concentrate on the parameters of interest without the need of knowing the particular pattern of the survival distribution.

In this chapter, an extended Cox proportional hazards model is proposed to approximate survival data in the presence of competing risks, covariates and immunes. Under the assumptions of independent censoring and improper baseline hazards, an MLE is derived from partial likelihood functions. The existence, consistency and asymptotic distributions of the estimators are studied thoroughly. A simulation is conducted to assess the performance of our proposed approach.

The rest of this chapter is organized as follows. In Section 5.2, we give the notations of the model and introduce some important relationships. In Section 5.3, we derive the estimators of the coefficients of covariates and failure rates using the partial likelihood (Cox, 1975) approach and establish their existence, consistency and asymptotic distributions. The analysis of the simulated data is described in Section 5.4. Section 5.5 gives the concluding remarks and discussions about further work. Finally, Section 5.6 is devoted to the mathematical proofs.

5.2 Specification

Like the models specified in Chapter 3 and Chapter 4, throughout this chapter, we assume that n individuals suffering from J risks are included in our study. Let T_i^* be an independent and continuous nonnegative random variable representing the true survival (uncensored) time of individual i, $i = 1, 2, \ldots, n$. Each individual i is observed over a limited time interval $[0, u_i]$, where u_i represents the censoring time which is independent of T_i^* and its distribution is assumed to be non-informative. Further let D_i be a discrete random variable with $D_i = j$ for $j = 1, 2, \ldots, J$ if the failure of the i-th individual will be caused by risk j and $D_i = 0$ implies that the failure will never be observed. Suppose that t_i^* is the realization of T_i^*, the *risk indicators* are $\delta_{ij} = 1_{\{t_i^* \le u_i, D_i = j\}}$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, J$ and the *censoring indicators* are $\delta_i = \sum_{j=1}^{J} \delta_{ij} = 1_{\{t_i^* \le u_i\}}$, $i = 1, 2, \ldots, n$, where $1_{\{E\}}$ is the indicator of event E. It is natural that the observation of individual i is comprised of (t_i, δ_{ij}, z_i), where $t_i = t_i^* \wedge u_i$ and z_i is an outcome of a κ-dimensional covariate vector Z_i which is independent of time.

The existence of immunes calls for a possibly improper hazard function including the informations of the instantaneous failure rates, the influence of covariates and the probability associated with the immunes. A common methodology adopted in the literature is to satisfy the requirements by letting the baseline hazard be defined by the real improper c.d.f. of T_i^* and the cured rate can be obtained by letting the survival time be ∞ in the estimated survival function. Instead we propose an approach in which the cured rate is included in the partial likelihood function and can be estimated directly.

The hazard functions are formulated by

$$h_{ij}(t)dt = p_j \exp\{\beta^T z_i\} h_0(t)dt = P\{t \leqslant T_i^* < t + dt, D_i = j | T_i^* \geqslant t\},$$

where $h_0(t)$ is an unspecified non-negative function which serves as the baseline hazard in our study, β is a κ-vector of unknown regression coefficients and $p_j = \exp(r_j)$ is the baseline probability of failing from cause j with $0 < p_j < 1$ and $\sum_{j=1}^{J} p_j = p \leq 1$.

Let $p_{ij} = P\{D_i = j\}$ be the probability that individual i dies (will die) from cause j and $F_{ij}(t) = P\{T_i^* \leq t | D_i = j\}$ represent the c.d.f. of the survival time with respect to a particular risk for $i = 1, \ldots, n$ and $j = 1, 2, \ldots, J$. Then $F_i(t)$, the c.d.f. of T_i^*, is

$$F_i(t) = P\{T_i^* \leq t\} = \sum_{j=1}^{J} P\{T_i^* \leq t | D_i = j\} P\{D_i = j\} = \sum_{j=1}^{J} p_{ij} F_{ij}(t).$$

It is obvious that $F_{ij}(t)$ is proper and $F_i(t)$ is proper if and only if $\sum_{j=1}^{J} p_{ij} = 1$.

Let $p^i = \sum_{j=1}^{J} p_{ij}$ be the probability that individual i is not an immune subject. From the definition of hazard functions we have:

$$1 - F_i(t) = \exp\left\{ - \int_{u=0}^{t} p \exp\{\beta^T z_i\} h_0(u) du \right\} = \{1 - F_0(t)\}^{\exp\{\beta^T z_i\}}, \tag{5.2.1}$$

where $F_0(t)$ is the baseline distribution function of T_i^*. This yields:

$$p^i = F_i^*(\infty) = 1 - (1 - p)^{\exp\{\beta^T z_i\}}. \tag{5.2.2}$$

Equation (5.2.2) shows how the probability of being an immune subject of an individual is influenced by the values of the covariates. Suppose that there is a baseline individual 0 with $Z_0 = 0$. Further note that

$$h_0(t)dt = \frac{P\{t \leqslant T_0^* < t + dt | T_0^* \geqslant t\}}{P\{D_0 \neq 0\}}. \tag{5.2.3}$$

It can be seen that $h_0(t)$ is the instaneous failure rate for individual 0 if and only if $P\{D_0 \neq 0\} = 1$, i.e. there are no immunes.

5.3 Likelihood and Asymptotics

In this section, we formulate our model on the basis of the previous specifications and derive the consistency and asymptotic normality of the estimators. The real survival time and the censoring time are assumed to be random on $[0, \tau_T]$ and $[0, \tau_C]$ with $0 \leq \tau_T, \tau_C \leq \infty$. Suppose that $\tau = \tau_T \wedge \tau_C$, the observation period is $[0, \tau]$.

As the baseline-hazard function $h_0(t)$ is left unspecified, we follow the approach of Cox's partial likelihood (Cox, 1975) with $h_0(t)$ left non-informative for the estimation. Let $\psi = (\beta^T, p_1, \ldots, p_J)^T$ be the parameter vector with true value $\psi_0 = ({\beta_0}^T, p_{10}, \ldots, p_{J0})^T$ and $\gamma = (\beta^T, r_1, \ldots, r_J)^T$, where $r_j = \log(p_j)$ for $1 \leq j \leq J$. The problem is to estimate the true value ψ_0 from the log likelihood function L on the basis of these observations and the estimator $\hat{\psi}$ is defined to be the solution of $\partial L(\psi)/\partial\psi = 0$. Noting the fact that $\hat{\gamma} = (\hat{\beta}^T, \log\hat{p_1}, \ldots, \log\hat{p_J})^T$ is exactly the solution of $\partial L(\gamma)/\partial\gamma = 0$ due to the good property of logarithm function, the problem of finding the estimator of ψ is equivalent to deriving the estimator of γ from the same likelihood function. To utilize the existing results of the literature, we take γ as the parameter vector to be estimated instead of ψ. Let $t_{(1)} < t_{(2)} < \cdots < t_{(K)}$ represent the ordered failure times. Further assume that there are no tied failure times and individual i_k is the the one who failed from risk j_k at $t_{(k)}$ for $k = 1, \ldots, K$. Denote by $R_i = R(t_i)$ the risk set (the individuals that are still under observation) at time t_i and set $r = \log p$, the partial likelihood function with respect to γ based on the observations (t_i, δ_{ij}, z_i) for $1 \leq i \leq n$

can be written as

$$L_p = \prod_{k=1}^{K} \frac{h_{i_k j_k}(t_{(k)})}{\sum_{i \in R_{i_k}} h_i(t_{(k)})} = \prod_{i=1}^{n} \frac{\prod_{j=1}^{J} \left(\exp(\beta^T z_i + r_j)\right)^{\delta_{ij}}}{\left(\sum_{l \in R_i} \exp(\beta^T z_l + r)\right)^{\delta_i}}, \tag{5.3.4}$$

where $h_i(t) = \sum_{j=1}^{J} h_{ij}(t)$ is the hazard function for individual i.

The process $N_{ij}(t) = I\{T_i^* \leq t, \delta_{ij} = 1\}$ is used to count the number of the j-th failures happening to individual i and $Y_i(t) = Y_{ij}(t) = I\{T_i \geq t\}$ indicates whether the ith individual is still under consideration for $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, J\}$. The sample paths of $N_{i1}, \ldots, N_{iJ}$ are step functions, zero at time zero, with at most one jump of size $+1$ at a time as there are no ties. Through this chapter, all properties of the counting processes of individual i are relative to the right continuous non-decreasing filtration $\{\mathscr{F}_i(t) : 0 \leq t \leq \tau\}$ with $\mathscr{F}_i(t) = \sigma\{N_{ij}(u), N_i^U(u) : 0 \leq u \leq t, j \in \{1, \ldots, J\}\}$, where $N_i^U(u) = I\{C_i \leq u, \delta_i = 0\}$. Obviously $N_i(t) = \sum_{j=1}^{J} N_{ij}(t)$ is also a counting process with at most one jump at a time and $Y_i(t)$ is a predictable process. Further we define

$$dN_{ij}(t) = N_{ij}(t) - N_{ij}(t-) \qquad \text{and} \qquad Q^{(l)}(\beta, t) = \sum_{i=1}^{n} Z_i^l Y_i(t) \exp\{\beta^T Z_i\}$$

for $l \in \{0, 1, 2\}$, where $Z_i^0 = 1$, $Z_i^1 = Z_i$ and $Z_i^2 = Z_i Z_i^T$.

Naturally, the intensity processes of $N_{ij}(t)$ and $N_i(t)$ conditional on the covariates are $\lambda_{ij}(t) = \exp\{\beta^T z_i + r_j\} Y_i(t) h_0(t)$ and $\lambda_i(t) = \exp\{\beta^T z_i + r\} Y_i(t) h_0(t)$ respectively. Thus

$$L_p = \prod_{i=1}^{n} \prod_{j=1}^{J} \prod_{t} \left\{ \frac{\exp\{\beta^T z_i + r_j\}}{\sum_{l=1}^{n} Y_l(t) \exp\{\beta^T z_l + r\}} \right\}^{dN_{ij}(t)}, \tag{5.3.5}$$

$$\log L_p = \sum_{i=1}^{n} \sum_{j=1}^{J} \int_{t=0}^{\infty} \left\{ \left(r_j + \beta^T z_i\right) - r - \log\left\{ \sum_{l=1}^{n} Y_l(t) \exp(\beta^T z_l) \right\} \right\} dN_{ij}(t). \quad (5.3.6)$$

Consider the parameter space $\Gamma = \left\{\gamma : \gamma = (\beta^T, r_1, \ldots, r_J)^T\right\}$ with the constrains that $\beta \in \mathbb{R}^\kappa, (r_1, \ldots, r_J)^T \in M^J$ and the true value $\gamma_0 = (\beta^T, r_{10}, \ldots, r_{J0})^T \in \Gamma$. The simplex M^J is defined by

$$M^J = \left\{ (r_1, \ldots, r_J)^T : r_j \in (-\infty, 0) \text{ for } j \in \{1, \ldots, J\}, \quad \sum_{j=1}^{J} \exp\{r_j\} \leq 1 \right\} \subset \mathbb{R}^J. \quad (5.3.7)$$

We aim to find a maximizer $\hat{\gamma} = (\hat{\beta}^T, \hat{r}_1, \ldots, \hat{r}_J)^T$ of $\log L_p$ and take it as the estimate of the true parameter γ_0. The following theorem shows the conditions for the existence of the consistent MLE and the asymptotic distributions of the estimators. We shall take advantage of the results in Andersen and Gill (1982) who provided a general approach of deriving the asymptotic properties of the Cox's regression model. They proved the consistency and asymptotic normality of the estimated covariate coefficients and suggested the asymptotic distribution of the baseline cumulative incidence function under the so called "A-D conditions" with the time scale being $[0, 1]$ in Section 3 of their paper. In addition, they also showed how the A-D conditions could be satisfied in i.i.d. case and the fact that time scale can be extended to $[0, \infty)$ under some additional conditions. As a result, our proof suffices to check the A-D conditions for $t \in [0, 1]$ as well as the condition of extending the result to $t \in [0, \infty)$.

Let e_j be the J-dimensional unit column vector with the jth element being 1 and

$Z_{ij} = (Z_i^T, e_j^T)^T$ be a new covariate vector. Further suppose that for $l \in \{0, 1, 2\}$,

$$S^{(l)}(\gamma, t) = \frac{1}{nJ} \sum_{i=1}^{n} \sum_{j=1}^{J} Z_{ij}^l Y_{ij}(t) \exp\{\gamma^T Z_{ij}\} \tag{5.3.8}$$

and $s^{(l)}(\gamma, t) = E[S^{(l)}(\gamma, t)]$ with $Z_{ij}^0 = 1$, $Z_{ij}^1 = Z_{ij}$ and $Z_{ij}^2 = Z_{ij} Z_{ij}^T$, then we have the following theorem.

Theorem 5.3.1 *Suppose that N_i, Y_i, Z_i are i.i.d. replicates of (N, Y, Z) for $l = 0$, 1, 2, let $q^{(l)}(\beta, t) = E[Q^{(l)}(\beta, t)]$. Assume the following conditions hold:*

(I) p_{j0}, $p_0 \in (0, 1)$, *$h_0(t)$ is non-negative finite and Z is bounded.*

(II) *The matrix Υ is positive definite, where Υ is defined by*

$$\Upsilon = \int_0^\infty \left(q^{(2)}(\beta_0, t) - \frac{q^{(1)}(\beta_0, t) \left(q^{(1)}(\beta_0, t) \right)^T}{\left(q^{(0)}(\beta_0, t) \right)^T} \right) h_0(t) dt.$$

Then with probability approaching one, there exists a local maximizer $\hat{\gamma}$ such that $\hat{\gamma} \xrightarrow{p} \gamma_0$ and $n^{1/2}(\hat{\gamma} - \gamma_0) \xrightarrow{d} N(0, \Sigma^{-1})$, where

$$\Sigma = \int_0^\infty \left(s^{(2)}(\gamma_0, t) - \frac{s^{(1)}(\gamma_0, t) \left(s^{(1)}(\gamma_0, t) \right)^T}{\left(s^{(0)}(\gamma_0, t) \right)^T} \right) h_0(t) dt. \tag{5.3.9}$$

Remark. Theorem 5.3.1 establishes the existence, consistency and asymptotic normality of $\hat{\gamma}$. So the estimated coefficients of the covariates have asymptotic normal distributions. The large sample distribution of the estimator of p_{j0} is log-normal instead of normal as a result of the assumption that $r_j = \log p_j$. Another issue to be noted is that our discussion is restricted to the interior of the parameter space. The boundary on which there are no immunes is not studied here.

Although we have restricted ourselves to the independence of baseline failure probabilities of the risks and Z_i, and the independence of β and failure causes, there is no difficulty in principle in allowing the cause-specific hazard function to be

$$h_{ij}(t)dt = p_j(z_i)\exp\{\beta_j^T z_i\}h_0(t)dt,$$

where $p_j(z_i)$ is a function of z_i and β_j is the regression coefficient associated with risk j. For instance, if we assume that

$$p_j(z_i) = \frac{\exp\{u_j + \pi_j^T z_i\}}{1 + \sum_{l=1}^{J}\exp\{u_l + \pi_l^T z_i\}},$$

the statistical inference can be given in the same way under some regular conditions. The assumption of proportional cause-specific hazards function, however, cannot be relaxed in this study. If we let $h_0 = h_{j0}$, i.e. $h_{ij}(t)dt = p_j\exp\{\beta^T z_i\}h_{j0}(t)dt$ such that the hazard functions are not proportional to each other, the partial likelihood approach cannot be adopted. The work of Larson and Dinse (1985) is relatively general and does not have such restrictions. But their work is based on a parametric setting and the exhaustion of the failure cause.

Following the suggestion of Breslow (1972), the cumulative baseline hazard can be estimated by

$$\hat{\Lambda}_0(t) = \int_0^t \frac{d\bar{N}(u)}{\sum_i \exp\{\hat{\beta}^T z_i\}Y_i(u)},$$

where $\Lambda_0(t) = \int_0^t p_0 h_0(u)du$ and $\bar{N}(u) = \sum_{i,j} N_{ij}(u)$. Consequently, the baseline proportion of the immunes can be estimated by

$$1 - \hat{p_0} = \exp\{-\hat{\Lambda}_0(\infty)\}$$

and the failure function of T_i^* can be approximated by

$$\hat{F}_i(t) = 1 - \hat{S}_i(t) = 1 - \exp\left\{-\exp(\hat{\beta}^T z_i)\hat{\Lambda}_0(t)\right\}. \tag{5.3.10}$$

The foregoing describes how the estimated probability of immunes may be derived with improper baseline c.d.f. mentioned in Section 5.2. The likelihood function employed is

$$\log L_p = \sum_{i=1}^{n} \int_0^\infty \beta^T z_i dN_i(t) - \int_0^\infty \log \left\{ \sum_{l=1}^{n} Y_l(t) \exp\{z_l^T \beta\} \right\} d\bar{N}(t). \tag{5.3.11}$$

It is obvious that only covariate is informative in Equation (5.3.11). The probabilities associated with the risks can be estimated by a non-parametric approach. In comparison, we have derived the MLE of p_{j0} for $1 \leq j \leq J$ parametrically.

5.4 A Simulation Study

A simulation has been conducted to assess the performance of our approach. Assume that $Z_i = (Z_{i1}, Z_{i2})$, where Z_{i1} follows a 50-50 Bernoulli distribution and $Z_{i2} \sim U(-1, 1)$. The survival times and failure causes of the individuals are assumed to be independent. The coefficient vector of the covariates is assumed to be β_0. Individual i suffers from two potential risks with associated probabilities p_{i1} and p_{i2} respectively. $p^i = p_{i1} + p_{i2}$ is the probability that individual i is not a long-term survivor, i.e. individual i will eventually fail from the two risks under consideration. We take $p^i < 1$ so that the immunes exist.

Without loss of generality, the baseline failure distribution given that a failure will eventually happen is assumed to be

$$F_0^*(t) = P\{T_0^* \leqslant t | D_0 \neq 0\} = 1 - e^{-0.0321t},$$

the baseline failure proportion for the two risks are set to be

$$p_{01} = P\{D_0 = 1\} = 0.4, \qquad p_{02} = P\{D_0 = 2\} = 0.5$$

and the covariate coefficient $\beta_0 = (-0.6484, 0.5)^T$. In addition, each subject is assumed to have an independent random censoring time generated from uniform distribution between 50 and 100. Equation 5.2.2 indicates that:

$$p^i = 1 - (1 - p^0)^{\exp\{\beta_0^T z_i\}}.$$

Thus

$$p_{i1} = p^i \, p_{01}/(p_{01} + p_{02}), \qquad p_{i2} = p^i \, p_{02}/(p_{01} + p_{02}).$$

The baseline distribution function is

$$\begin{aligned} F_0(t) &= P\{T_0^* \leqslant t\} \\ &= P\{T_0^* \leqslant t, D_0 \neq 0\} \\ &= P\{T_0^* \leqslant t | D_0 \neq 0\}(p_{01} + p_{02}) \\ &= 0.9 - 0.9e^{-0.0321t}. \end{aligned}$$

Hence the true distribution function for T_i^* is

$$\begin{aligned} F_i(t) &= P\{T_i^* \leqslant t\} \\ &= 1 - P\{T_i^* > t\} \\ &= 1 - \big(1 - F_0(t)\big)^{\exp\{\beta_0^T z_i\}} \\ &= 1 - (0.1 + 0.9e^{-0.0321t})^{\exp\{\beta_0^T z_i\}}. \end{aligned}$$

Similarly we can derive the distribution conditional on the eventual failure. It is

$$F_i^*(t) = \frac{F_i(t)}{p_{i1} + p_{i2}} = \frac{1 - (0.1 + 0.9e^{-0.0321t})^{\exp\{\beta_0^T z_i\}}}{1 - (1 - p^0)^{\exp\{\beta_0^T z_i\}}}.$$

Table 1 shows the simulation performance for different sample sizes with replicates $R = 100$. To show the robustness of this approach, another simulation with $\beta_0 = (-0.2730, 0.8)^T$, $p_{01} = 0.3$ and $p_{02} = 0.4$ is also conducted. The results have been listed in Table 2 also with replicates $R = 100$. It can be seen from Table 5.1 and Table 5.2 that our proposed approach leads to reasonably accurate point estimates. The STD also seems to decrease as the sample size increases.

In addition, to assess the goodness-of-fit of our fitted model, Kolmogorov-Smirnov test has been carried out for different sample sizes of 1000 replicates based on the removal of the continuous covariate. Two groups of survival data are generated separately with the covariates $(0, 0)^T$ and $(1, 0)^T$ for the analysis. The sample size for each group is assumed to be M. As can be seen from Table 5.3, the rejection percentage for each sample size is almost zero.

Figures 5.1 and 5.2 show the curves of the estimated c.d.f.'s for the grouped data with covariates $(0, 0)^T$ and $(1, 0)^T$ superimposed by the empirical distribution and true distribution for one realization of sample size 100 for each group in simulation 1. The empirical distribution functions are derived by Kaplan-Meier method (Kaplan and Meier, 1958) and the fitted c.d.f.'s are developed from Equation (5.3.10). It can be seen that the curves are similar and the proposed approach provides good descriptions of the c.d.f.'s for both groups.

Table 5.1: Estimated Parameters in Simulation 1

	Parameter	p_{01}	p_{02}	$\beta_0(1)$	$\beta_0(2)$
N	True value	0.4000	0.5000	-0.6484	0.5000
100	Estimate	0.3989	0.4951	-0.6653	0.5261
	STD	0.1153	0.1361	0.2862	0.2201
200	Estimate	0.3983	0.4914	-0.6676	0.5327
	STD	0.0742	0.0919	0.1949	0.1478
300	Estimate	0.4045	0.4998	-0.6614	0.5213
	STD	0.0594	0.0639	0.1507	0.1337
400	Estimate	0.4022	0.4966	-0.6655	0.5186
	STD	0.0497	0.0533	0.1350	0.1110
500	Estimate	0.4010	0.4984	-0.6596	0.5201
	STD	0.0453	0.0517	0.1172	0.1046
600	Estimate	0.4026	0.5009	-0.6588	0.5138
	STD	0.0417	0.0444	0.0959	0.0989
700	Estimate	0.3972	0.4958	-0.6590	0.5132
	STD	0.0364	0.0397	0.0832	0.0851
800	Estimate	0.3973	0.4991	-0.6563	0.5163
	STD	0.0339	0.0439	0.0860	0.0845
900	Estimate	0.3998	0.5012	-0.6576	0.5122
	STD	0.0314	0.0383	0.0769	0.0810
1000	Estimate	0.4014	0.5010	-0.6521	0.5106
	STD	0.0302	0.0356	0.0735	0.0725

Table 5.2: Estimated Parameters in Simulation 2

	Parameter	p_{01}	p_{02}	$\beta_0(1)$	$\beta_0(2)$
N	True value	0.3000	0.4000	-0.2730	0.8000
100	Estimate	0.2936	0.4068	-0.2914	0.7827
	STD	0.1097	0.1693	0.2687	0.2478
200	Estimate	0.2969	0.4038	-0.2848	0.8062
	STD	0.0984	0.1386	0.2104	0.1665
300	Estimate	0.3001	0.4072	-0.2842	0.7951
	STD	0.0643	0.0976	0.1656	0.1307
400	Estimate	0.3006	0.4087	-0.2889	0.7986
	STD	0.0672	0.0917	0.1411	0.1149
500	Estimate	0.3055	0.4138	-0.2812	0.7960
	STD	0.0509	0.0748	0.1282	0.1142
600	Estimate	0.3041	0.4067	-0.2846	0.7919
	STD	0.0513	0.0707	0.1100	0.0955
700	Estimate	0.3005	0.4016	-0.2786	0.7919
	STD	0.0406	0.0633	0.1005	0.0891
800	Estimate	0.3019	0.4031	-0.2792	0.7901
	STD	0.0388	0.0544	0.0964	0.0868
900	Estimate	0.3033	0.4031	-0.2773	0.7904
	STD	0.0369	0.0540	0.0882	0.0746
1000	Estimate	0.3002	0.4010	-0.2753	0.7924
	STD	0.0337	0.0473	0.0872	0.0731

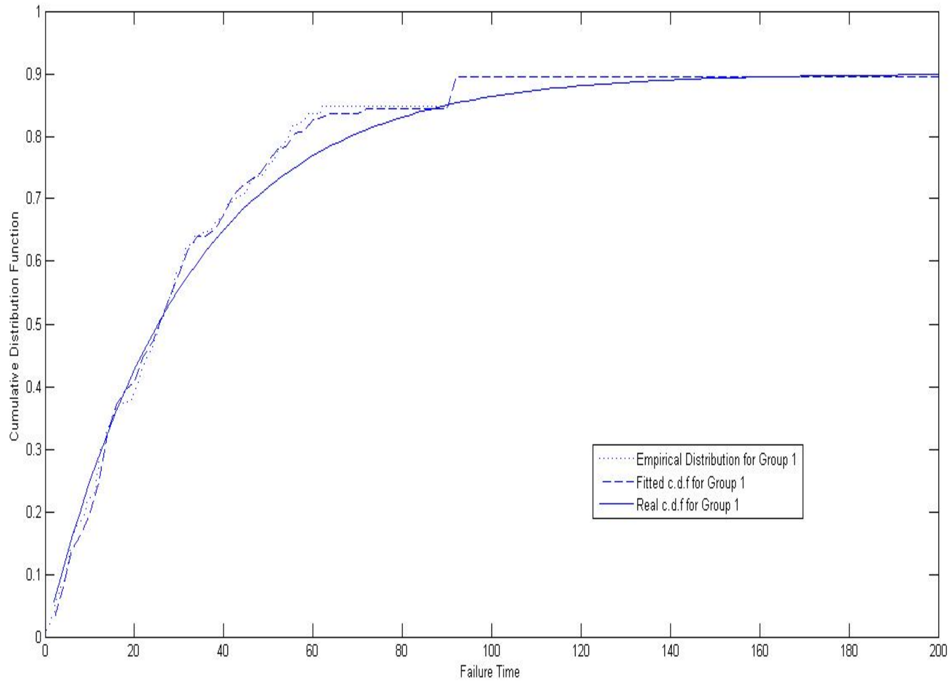

Figure 5.1: Curves of C.d.f.'s for Group 1

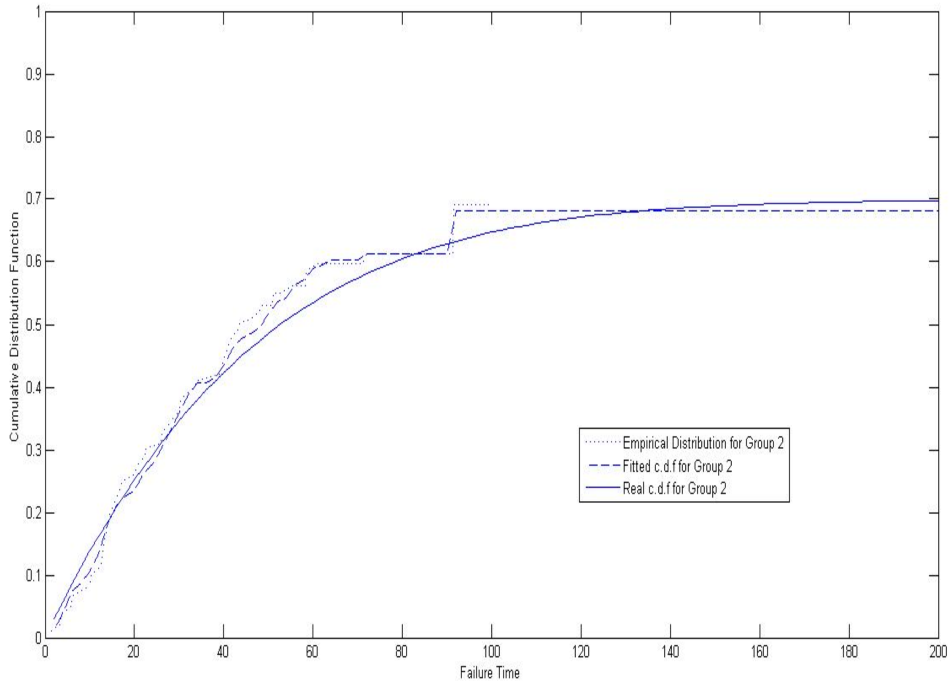

Figure 5.2: Curves of C.d.f.'s for Group 2

Table 5.3: Rejection Percentage by K-S Test

	M	20	50	100	200	300	400	500
Simulation 1	Group 1	0.009	0.012	0.006	0.004	0.006	0.007	0.005
	Group 2	0.018	0.019	0.013	0.016	0.012	0.014	0.011
Simulation 2	Group 1	0.011	0.008	0.014	0.012	0.009	0.004	0.003
	Group 2	0.013	0.008	0.007	0.009	0.008	0.007	0.008

5.5 Concluding Remarks

An extended Cox proportional hazards model is proposed to approximate survival data for competing risks in the presence of covariates and immunes. Under the assumptions of independent censoring and improper baseline hazards, an MLE can be derived from the partial likelihood function. The existence, consistency and asymptotic properties of the estimators are studied thoroughly. In particular, the estimators of the failure rates are proved to be consistent and asymptotically log-normally distributed. A simulation study which favors the proposed approach has been conducted to assess and compare the performances of our methods under different sample sizes.

The advantage of this work is that the probabilities of failing from the risks are parameterized and separated from the shape of baseline hazard functions. In previous work, the probabilities can only be obtained after the hazard functions are estimated first. In many studies, the hazard functions are not of main interest. Our approach provides an alternative way to investigate these problems. At least, it provides a method to verify the results developed by the traditional approach. In addition, it also provides a foundation of the study of the existence of immunes in a semi-parametric approach.

We have restricted our discussion in the proportional hazards setting which is expected to be relaxed in the future based on the development of counting processes techniques. Useful approaches such as those of Aalen (1975), Kuk and Chen (1992) and Andersen et al. (1993) have provided a basis for this development. Another interesting issue is the test for the existence of immunes, i.e. whether the value of $\sum_{j=1}^{J} p_{0j}$ is 1 or not. A boundary test of "no immunes" for competing risks data in the presence of covariates has been discussed in Chapter 3 and a similar test in the framework of semi-parametric model is also included in the future work.

5.6 Proofs

The proof will make use of to the work of Andersen and Gill (1982), where a general approach to the large sample properties of Cox's regression model is adopted. Their conditions A-D can be recast as:

A. $\int_0^1 h_0(t)dt < \infty$.

B. There exists a neighborhood Γ of γ_0 such that for $l = 0, 1, 2$

$$\sup_{t\in[0,1],\gamma\in\Gamma} ||S^{(l)}(\gamma, t) - s^{(l)}(\gamma, t)|| \xrightarrow{P} 0.$$

C. There exists $\delta > 0$ such that

$$n^{-1/2} \sup_{i,j,t} |Z_{ij}| Y_{ij}(t) I\{\gamma_0^T Z_{ij} > -\delta\} |Z_{ij}| \xrightarrow{P} 0.$$

D. $s^{(l)}(\gamma, t)$ is continuous and bounded in Γ uniformly in $t \in [0,1]$ for $l = 0,1,2$ with $s^{(0)}$ away from zero.

Lemma 5.6.1 *(The application of Theorem 4.1 and Theorem 4.2 of Andersen and Gill (1982)) Suppose that N_i, Y_i, Z_i are i.i.d. replicates of (N, Y, Z) with Y left continuous with right hand limits, Z is bounded, $E[N(\infty)] < \infty$, the above A-D conditions are satisfied and for each $\tau < \infty$, $P\big(Y(t) = 1, \forall t < \tau\big) > 0$, we have:*

$$n^{1/2}(\hat{\gamma} - \gamma_0) \xrightarrow{D} \mathcal{N}(0, \Sigma^{-1}). \tag{5.6.12}$$

Proof. The proof is standard and omitted here. ■

Proof of Theorem 5.3.1 Note that $Y(t)$ is automatically left continuous with right hand limits and $E[N(\infty)] < \infty$. The log partial likelihood function can be rewritten as

$$\log L_p = \sum_{i,j} \int_{t=0}^{\infty} z_{ij}^T \gamma dN_{ij}(t) - \sum_{i,j} \int_{t=0}^{\infty} \log \left\{ \sum_{l,m} Y_{lm}(t) \exp\{z_{lm}^T \gamma\} \right\} dN_{ij}(t). \tag{5.6.13}$$

The score function with respect to γ is

$$U(\gamma) = \sum_{i,j} \int_0^{\infty} \left\{ z_{ij} - \frac{\sum_{l,m} Y_{lm}(t) z_{lm} \exp\{z_{lm}^T \gamma\}}{\sum_{l,m} Y_{lm}(t) \exp\{z_{lm}^T \gamma\}} \right\} dN_{ij}(t) \tag{5.6.14}$$

with a compensator

$$C(\gamma) = \sum_{i,j} \int_0^\infty \left\{ z_{ij} - \frac{\sum_{l,m} Y_{lm}(t) z_{lm} \exp\{z_{lm}^T \gamma\}}{\sum_{l,m} Y_{lm}(t) \exp\{z_{lm}^T \gamma\}} \right\} dA_{ij}(t), \tag{5.6.15}$$

where

$$A_{ij}(t, \gamma) = \int_0^t \lambda_{ij}(t) dt \tag{5.6.16}$$

is the compensator of $N_{ij}(t)$.

The next part of the proof is to check the A-D conditions similar to that in Andersen and Gill (1982). Note that,

$$S^{(0)}(\gamma, t) = \frac{1}{nJ} \sum_{i=1}^{n} \sum_{j=1}^{J} Y_{ij} \exp^{\gamma^T Z_{ij}(t)} = \frac{p}{nJ} \sum_{i=1}^{n} Y_i(t) \exp^{\beta^T Z_i(t)}, \tag{5.6.17}$$

$$S^{(1)}(\gamma, t) = \frac{1}{nJ} \sum_{i=1}^{n} \sum_{j=1}^{J} Z_{ij}(t) Y_{ij} \exp^{\gamma^T Z_{ij}(t)} = \frac{p}{nJ} \sum_{i=1}^{n} Y_i(t) \exp^{\beta^T Z_i(t)} (Z_i^T, w^T)^T, \tag{5.6.18}$$

$$S^{(2)}(\gamma, t) = \frac{1}{nJ} \sum_{i=1}^{n} \sum_{j=1}^{J} Z_{ij} Z_{ij}^T(t) Y_{ij} \exp^{\gamma^T Z_{ij}(t)} = \frac{p}{nJ} \sum_{i=1}^{n} Y_i(t) \exp^{\beta^T Z_i(t)} \Delta \tag{5.6.19}$$

and $s^{(i)}(\gamma, t) = E[S^{(i)}(\gamma, t)]$, where the vector $w = (p_1, \ldots, p_J)^T / p$ is comprised of the weights of the risks and

$$\Delta = \begin{pmatrix} Z_i Z_i^T & Z_i w^T \\ w Z_i^T & w I_J \end{pmatrix}. \tag{5.6.20}$$

After some trivial calculation, we can see that

$$\Sigma = \begin{pmatrix} \Upsilon & 0 \\ \\ \Psi & \Xi \end{pmatrix} \text{ with } \Xi = \int_0^\infty q_0(\beta_0, t) w(I_J - wI_J) h_0(t) dt\,.$$

Ψ is a $\kappa \times J$ matrix and its value has no influence on the result. Obviously, Σ is positive definite given that Υ and Ξ are positive definite by assumption. Further for each $t < \tau_G \leq \infty$, we have

$$P\{Y_i(t) = 1\} = P\{T_i^* \geq t, u_i \geq t\} \geq (1-p)(1 - G(t-)) > 0\,, \tag{5.6.21}$$

where G is the c.d.f. of u_i. Therefore by the same line of Theorems 4.1 and 4.2 of Andersen and Gill (1982) the conditions A, B and D hold. This leaves the condition "C". Note that

$$\begin{aligned}
& n^{-1/2} \sup_{i,j,t} |Z_{ij}(t)| Y_{ij}(t) I\{\gamma_0^T Z_{ij}(t) > -\delta |Z_{ij}(t)|\} \\
&= n^{-1/2} \sup_{i,j,t} \sqrt{Z_i(t)^2 + 1} Y_i(t) I\left\{\beta_0^T Z_i(t) + \log p_j > -\delta \sqrt{Z_i(t)^2 + 1}\right\} \\
&\leq n^{-1/2} \sup_{i,j,t} |Z_i(t)| Y_i(t) I\left\{\beta_0^T Z_i(t) > -\delta \sqrt{Z_i(t)^2 + 1}\right\} + n^{-1/2} \\
&= B + n^{-1/2}.
\end{aligned}$$

It is obvious that $B \xrightarrow{P} 0$. Condition "C" holds consequently. To extend the time scale, Z is bounded by previous specifications, $E[N(\infty)] \leq 1 < \infty$ and for each $\tau < \infty$, $P(Y(t) = 1, \forall t < \tau) > 0$ by Equation (5.6.21).

The proof is completed.

Reference

Aalen, O. O. (1975). *Statistical inference for a family of counting processes.* Ph.D. Thesis, University of California, Berkeley.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.*, **10**, 1100-1120.

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

Bernoulli, D. (1760). Essai d'une nouvelle analyse de la mortalité causée par la petite Vérole, et des avantages de l'Inoculation pour la prévenir, *Mém. Acad. Roy. Sci.*, 1-45.

Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J. Roy. Statist. Soc. Ser. B*, **11**, 15-45.

Breslow, N. (1972). Contribution to the discussion on the paper of Cox (1972).

Choi, K. C. and Zhou, X. (2002). Large sample properties of mixture models with covariates for competing risks. *J. Multivariate Anal.*, **82**, 331-336.

Cox, D. R. (1959). The analysis of exponentially distributed life-times with two types of failure. *J. Roy. Statist. Soc. Ser. B*, **21**, 411-421.

Cox, D. R. (1970). *The analysis of binary data.* London: Methuen.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B*, **34**, 187-220.

Cox, D. R. (1975). Partial Likelihood. *Biometrika*, **62**, 269-276.

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

Crowley, J and Hu, M. (1977). Covariance Analysis of Heart Transplant Survival Data. *J. Amer. Statist. Ass.*, **72** 27-36.

Craiu, R. V. and Duchesne, T. (2004). Inference based on the EM algorithm for the competing risks model with masked causes of failure. *Biometrika*, **91**, 543-558.

Fleming, T. R. and Harrington, D. P. (1991). *Counting processes and survival analysis.* New York: Wiley.

Friedman, M. (1982). Piecewise Exponential Models for Survival Data with Covariates Michael. *Ann. Statist.*, **10**, 101-113.

Ghitany, M. E. and Maller, R. A. (1992). Asymptotic results for exponential mixture models with long term survivors. *Statistics*, **23**, 321-336.

Ghitany, M. E., Maller, R. A. and Zhou, S. (1994). Exponential mixture models with long-term survivors and covariates. *J. Multivariate Anal.*, **49**, 218- 241.

Han, A. and Hausman, J. A (1990). Flexible Parametric Estimation of Duration and Competing Risk Models. *J. Appl. Econometrics* , **5**, No. 1, 1-28.

Herman, R. J. and Patell, K. N. (1971). Maximum likelihood estimation for multi-risk model. *Technometrics*, **13**, No. 2, 385-96.

Ibrahim, J. G., Chen, M. and Kim, S. (2008). Bayesian Variable Selection For The Cox Regression Model With Missing Covariates. *Lifetime Data Anal.*, **14**, No. 4, 496-520.

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

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Ass.*, **53**, 457-481.

Kundu, D. and Basu, S. (2000). Analysis of incomplete data in presence of competing risks. *J. Statist. Plann. Infer.,* **87**, 221-239.

Kuk, Y. C. and Chen, H. C.(1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, **79**, 531-541.

Larson, M. G. and Dinse, G. E. (1985). A mixture model for the regression analysis of competing risks data. *Appl. Statist.*, **34**, 201-211.

Lagakos, S. W., Sommer, C. J. and Zelen, M. (1978). Semi-Markov models for partially censored data. *Biometrika*, **65**, 311-317.

Maller, R. A. and Zhou, X. (1996). *Survival Analysis with Long-term Survivors.* Chichester; New York: Wiley.

Maller, R. A. and Zhou, X. (2002). Analysis of parametric models for competing risks. *Statistica Sinica*, **50**, No. 3, 725-750.

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T. and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks.

Biometrics, **34**, 541-554.

Miyakawa, M. (1984). Analysis of incomplete data in a competing risks model. *IEEE Trans. Reliab.*, **33**, 293-296.

Peng, Y. and Dear, K. B. G. (2000). A Nonparametric Mixture Model for Cure Rate Estimation. *Biometrics*, **56**, 237-243.

Shu, Y. and Klein, J. P. (2005). Additive hazards Markov regression models illustrated with bone marrow transplant data. *Biometrika*, **92**, 283-301.

Vu, H. T. V., Maller, R. A. and Zhou, X. (1998). Asymptotic Properties of a Class of Mixture Models for Failure Data: The Interior and Boundary Cases. *Ann. Inst. Statist. Math.*, **50**, 627-653.

Vu, H. T. V. and Zhou, X. (1997). Generalization of likelihood ratio tests under non-standard conditions. *Ann. Statist.*, **25**, 897-916.

Wu, C. F. J. (1983). On the Convergence Properties of the EM Algorithm. *Ann. Statist.*, **11**, No. 1, 95-103.

Zhao, X. B. and Zhou, X. (2006). Proportional hazards models for survival data with long-term survivors. *Statist. Probab. Lett.*, **76**, 1685-1993.

Zhou, X. and Maller, R. A. (1995). The likelihood ratio test for the presence of immunes in a censored sample. *Statistics*, **27**, 181-201.