Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par
Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Discipline ou spécialité :
MATHEMATIOUES APPLIOUEES

Présentée et soutenue par
Angelica HERNANDEZ QUINTERO

Le *vendredi 11 juin 2010*

Titre :
Inférence statistique basée sur les processus empiriques dans des modèles semi-paramétriques de durées de vie□□□□□

JURY
DAUXOIS Jean-Yves (Université de Besançon, Rapporteur)
NIETO BARAJAS Enrique (Instituto Tecnológico Autónomo, Mexico, Rapporteur)
GAMBOA Fabrice (Université Paul Sabatier – Toulouse 3, Examinatuer)
GORDIENKO ILLICH Evgueni (Universidad Autónoma Metropolitana, Mexico, Examinatuer)
MONTES DE OCA MACHORRO José Raúl(Universidad Autónoma Metropolitana, Mexico, Examinatuer)
DUPUY Jean-François (Université de La Rochelle, Directeur de thèse)

Ecole doctorale : *Mathématiques Informatique Télécommunications (MITT)*
Unité de recherche : *Institut de Mathématiques de Toulouse*
Directeur(s) de Thèse : *DUPUY Jean-François et ESCARELA Gabriel*
Rapporteurs : *DAUXOIS Jean-Yves et NIETO BARAJAS Enrique*

Remerciements

Je tiens tout d'abord à remercier Gabriel et Jean-François, qui m'avais permis de travailler avec vous. Merci pour me donner votre soutien, en particulier, pour la confiance que vous m'avez témoignée en me proposant ce sujet de thèse, sans vous cette thse n'aurait pu être possible. Merci pour votre temps, votre dévouement et vos conseils qui ont permis ce travail.

Je tiens à remercier Jean-Yves Dauxois, Fabrice Gamboa, Evgueni Gordienko, Raul Montes de Oca et Luis Enrique Nieto, qui m'a fait l'honneur de présider le jury. Je leur suis particulièrement reconnaissant pour le nombreux commentaires lesquels ont permis d'enrichir et de compléter cette thèse.

Merci Marco pour ton soutien et votre compréhension. Et surtout pour toute l'aide et des conseils que tu m'as donné pour développer la thèse. Je remercie également à ma famille qui a toujours été là.

À Marco par son soutien et pour que tu as toujours été là,

À ma famille pour être présent à chaque instant.

Summary

Survival data arise from disciplines such as medicine, criminology, finance and engineering amongst others. In many circumstances the event of interest can be classified in several causes of death or failure and in some others the event can only be observed for a proportion of "susceptibles". Data for these two cases are known as competing risks and long-term survivors, respectively. Issues relevant to the analysis of these two types of data include basic properties such as the parameters estimation, existence, consistency and asymptotic normality of the estimators, and their efficiency when they follow a semiparametric structure. The present thesis investigates these properties in well established semiparametric formulations for the analysis of both competing risks and long-term survivors. It presents an overview of mathematical tools that allow for the study of these basic properties and describes how the modern theory of empirical processes and the theory of semiparametric efficiency facilitate relevant proofs. Also, consistent variance estimate for both the parametric and semiparametric components for the two models are presented.

The findings of this research provide the theoretical basis for obtaining inferences with large samples, the calculation of confidence bands and hypothesis testing. The methods are illustrated with data bases generated through simulations.

Keywords. Competing risks, empirical process, long-term survivors, mixture model, proportional hazard model, transformation model.

Résumé

L'analyse statistique de durées de vie censurées intervient dans de nombreuses disciplines, comme la médecine, la fiabilité, la criminologie, la finance, l'ingénierie. Chacun de ces domaines fournit des exemples de situations où: i) l'évènement observé est dû à une cause parmi plusieurs causes en compétition, ii) l'évènement ne peut être observé que pour une fraction, inconnue de l'analyste, de sujets "susceptibles". On parle respectivement de durées de vie en présence de risques concurrents, et de durées de vie en présence d'une fraction immune. Les problèmes posés pour l'analyse statistique de modèles de durées en présence de ces deux types de données incluent la construction d'estimateurs, l'étude de leurs propriétés asymptotiques (consistance, normalité asymptotique, efficacité, estimation de la variance asymptotique), et leur implémentation. Dans ce travail, nous nous intéressons à ce type de problèmes pour deux modèles de régression semi-paramétriques de durées de vie. Nous considérons successivement un modèle de mélange semi-paramétrique basé sur le modèle à risques proportionnels de Cox, puis le modèle de régression semi-paramétrique de transformation linéaire pour l'étude de durées de vie en présence d'une fraction immune. Nous construisons des estimateurs et établissons leurs propriétés asymptotiques, en utilisant des outils issus de la théorie des processus empiriques. Des études de simulation sont également menées.

Mots clés. Durées censurées, fraction immune, modèle de mélange, modèle à risques proportionnels, modèle de transformation linéaire, processus empiriques, propriétés asymptotiques, risques concurrents.

Contents

Introduction

Survival analysis is a branch of statistics which deals with a set of techniques and procedures to analyse data where the interest is to analyze the time between an initial event and a final event. These procedures represent nowadays a fundamental part in clinical trials, epidemiological studies, economics, actuarial science and engineering and many other disciplines. The main features of survival data are that the underlying distribution tends to have a long tail, there is a significant number of censored observations, which occur when survival times are the not exactly known and the data corresponds to the point where the experimental unit was seen for the last time alive.

The main interest in survival analysis is to estimate the survival function, which is the probability of experiencing an event after a certain period. Actually, there is well-established theory to estimate the survival function in the nonparametric, semiparametric and fully parametric form. In particular, the semiparametric Cox model represents a flexible way to model the survival function in the presence of explanatory variables as its proportional hazards structure allows to ignore the functional form of the baseline survival function and therefore it is possible to perform inferences about the parameters that are associated with the explanatory variables; in this case a modified likelihood function is used, which has very similar qualities to a completely specified likelihood function.

In the context of survival data, we can have that the occurrence of the event can be classified into different causes and further, such occurrence is caused by only one cause in particular. This type of data are known as *competing risks*. One of the main objectives of the study of these data is to model and estimate the cumulative incidence function of a specific cause type, which is the probability of occurrence of the event for some cause within a certain period, in the presence of explanatory variables. One way to deal with estimation is to regard individuals who do not experience the event of interest as censored. However, a patient experiencing a competing risk event is censored in informative form that is, he can not be excluded the study and, therefore, the methods used to analyze a single event are not helpful. In this case, the cumulative incidence function for an event of interest should be calculated taking into account the presence of other competing causes. There are several models for analyzing competing risks. Amongst them, the formulation proposed by Larson and Dinse (1985) stands out for its identifiability and easy interpreta-

tion. This model specifies the cumulative incidence functions in terms of probabilities of conditional survival of cause-specific and the probabilities that the event eventually occurs from a cause. Larson and Dinse proposed a fully parametric structure where the conditional survival functions have the form of the parametric Cox proportional-hazards whose baseline hazard function is piece-wise exponential and the probability of cause-specific follow a multinomial model. Recently, Ng and McLachlan (2003) and Escarela and Bowater (2008) have proposed a semiparametric extension of the formulation presented by Larson and Dinse in such a way that the conditional baseline survival functions are expressed through the semiparametric Cox proportional hazards model. Thus the resulting model provides a flexible and parsimonious way to study competing risks. These studies focus on the computational aspects of the estimation in the semiparametric mixture model but they have failed to analyse the large-sample properties of the estimators proposed by them.

It is common to find competing risks data where a risk is not observable and corresponds to a subgroup *immune* to the event of interest. For this type of data the graph of the empirical survival function shows a plateau, suggesting the existence of a proportion of individuals who will never experience the event of interest. An example of this type of data is when in clinical studies, a proportion of individuals respond favorably to treatment and subsequently appear to be free of any signs or symptoms of the disease and may be considered cured, while the remaining patients eventually experience relapse. Berkson and Gage (1952) used a mixture exponential distributions and a constant cure fraction to fit survival data from studies of breast cancer and stomach cancer. Kuk and Chen (1992) considered estimation of regression parameters using marginal likelihood method and proposed the so-called proportional hazards cure model in which the proportional hazards regression models (Cox, 1972) is specified in the survival times of susceptible subjects while the logistic regression models is utilized in the cure fraction. Peng and Dear (2000) and Sy and Taylor (2000) have proposed a nonparametric mixture model in which the assumption of proportional hazards is used for modeling covariate effects in times of failure from individuals who are not cured and propose an estimation method based on the EM algorithm. Recently, a generalization of these models has been proposed by Lu and Ying (2004), which employs transformation models to specify the time of failure of susceptible individuals and a logistic model to model the curable fraction.

The purpose of this thesis is to study the asymptotic properties of two classes of semiparametric regression models in survival analysis: a semiparametric mixture model for competing risks and the semiparametric transformation cure model. The first class corresponds to the formulation proposed by Escarela and Bowater (2008) whose formulation specifies separately the probability of eventually experience a type of failure and the conditional risk for each type of failure. The second model is a generalization the model proposed by Lu and Ying (2004) which has the flexibility to include effects of time-dependent covariates and combines a logistic regression for the probability of eventual occurrence with the class of transformation models for the times of occurrence. This generalization extends several cure models established in the literature such as the proportional hazards

model (Farewell, 1982; Kuk and Chen, 1992, Sy and Taylor, 2000, Peng and Dear, 2000) and the odd-proportional model (Lu and Ying, 2004).

The present work studies the asymptotic properties of nonparametric maximum likelihood estimators of the two classes of models on display. The method of nonparametric maximum likelihood estimation (NPML) is reviewed and it is shown that the formulation of this models lends itself to the use of tools from the modern theory of empirical processes to proof the asymptotic properties of the resulting estimators.

The two classes of models described above have infinite-dimensional parameters, which raises theoretical challenges for statistical analysis. In this thesis, the techniques developed by Murphy (1994.1995) and Parner (1998) for frailty models are extended and applied using the modern theory of empirical processes established by Van der Vaart and Wellner (1996), which makes it covenient to study the existence, consistency and normality asymptotic of the resulting estimates from both classes; moreover, it is shown that the maximum likelihood estimators for the regression parameters are semiparametric efficient (Bickel et al., 1993). Finally, we consistent variance estimators are proposed for both the finite and infinite dimensional parameters in these models.

The thesis is split into two parts. The first part comprises four chapters and gives a brief introduction to topics that will be useful for development of the work. Chapter 1 summarises basic definitions of survival analysis, some important survival parametric functions and the notation of right censoring. Also, both the parametric inference in presence of censoring data and the Cox model are reviewed. Chapter 2 describes definitions and properties essentials of counting processes and martingales. Chapter 3 outlines general semiparametric inference techniques, emphasizing the semiparametric efficiency. Chapter 4 is focused to explaining the theory of empirical processes which is a fundamental part in the study of asymptotic properties of the estimates from each kind of model studied here. The second part consists of two chapters in which the two classes of models described above are analysed. In Chapter 5 the semiparametric mixture model for competing risks proposed by Escarela and Bowater (2008) is reviewed including regularity conditions and the approach of maximum likelihood. The empirical process theory is applied to show consistency and asymptotic normality of the estimated maximum likelihood parameters. Consistent variance estimator are finally obtained. Finally, two simulated data sets are used to compare the fit of the semiparametric mixture model for competing risks with that of a parametric model. In Chapter 6 we discuss the semiparametric cure model constructed with transformation models. We give a brief introduction to cure models and linear transformation models. A specification of the cure model using transformation models, which allows for the inclusion of time-dependent variables, is given. It is then shown the existence, consistency, normality asymptotic and efficiency of maximum likelihood estimators nonparametric. Finally, a consistent estimator of the variance for the regression parameter and for the infinite-dimensional parameter is presented.

Chapter 7 presents conclusions and perspectives derived from the study of both classes of semiparametric models.

Introduction

L'analyse des durées de vie est un domaine de la statistique qui comprend un ensemble de techniques et de procédures pour analyser des données où la variable réponse est le temps écoulé entre un évènement initial et un évènement final. Ces techniques sont fondamentales pour l'analyse statistique des essais cliniques, des études épidémiologiques, et dans de nombreuses autres disciplines, comme l'économie, la science actuarielle, l'ingénierie. Une caractéristique fondamentale des durées de vie est la présence de censure, qui en complique l'analyse statistique.

Un objectif du statisticien confronté à des durées de vie consiste à estimer le modèle sous-jacent à ces données (en estimant, par exemple, une fonction de survie, une fonction de risque instantané ou cumulé, ou un paramètre de régression). Il existe des outils bien établis pour estimer de telles quantités, que ce soit dans un cadre paramétrique, semi-paramétrique, ou non-paramétrique. En particulier, le modèle semi-paramétrique de Cox représente un moyen flexible pour modéliser la fonction de survie en présence de variables explicatives. Des modèles plus généraux, tels que le modèle de transformation linéaire, suscitent actuellement un fort intérêt.

En analyse statistique des durées de vie, nous disposons parfois, en plus d'une durée observée, de la cause de l'évènement associé. Ces données interviennent en particulier dans un contexte de risques concurrents. L'un des objectifs principaux de l'étude de telles données, est de modéliser et estimer la fonction d'incidence cumulée d'une cause d'évènement spécifique. Il existe de nombreux modèles pour analyser des risques concurrents. Parmi ceux-ci, le modèle proposé et implémenté par Larson et Dinse (1985) se distingue par l'interprétation aisée des résultats qu'il produit. Ce modèle spécifie les fonctions d'incidence cumulée en terme des probabilités conditionnelles de survie sachant la cause spécifique d'évènement, et des probabilités de chaque cause spécifique. Larson et Dinse (1985) proposent un modèle complètement paramétrique, où les fonctions de survie sont spécifiées par des modèles à risques proportionnels de Cox dont la fonction de risque de base est constante par morceaux, et la loi des causes spécifiques est multinomiale. Récemment, Ng et McLachlan (2003) et Escarela et Bowater (2008) ont proposé une formulation semi-paramétrique du modèle de Larson et Dinse (1985), où les fonctions de survie conditionnelles aux régresseurs sont exprimées au travers du modèle

semi-paramétrique de Cox. Le modèle qui en résulte fournit une formulation flexible et parcimonieuse pour étudier des risques concurrents. Ces auteurs ont porté leur attention sur les aspects algorithmiques de l'estimation de ce nouveau modèle, en ignorant l'étude des propriétés asymptotiques des estimateurs proposés.

Il est également fréquent de rencontrer des données de survie où l'évènement n'est pas observable pour un groupe de sujets préservé du risque de survenue de cet évènement. Un exemple de ce type de données intervient dans les essais cliniques, quand une proportion d'individus qui a répondu favorablement à un traitement cesse d'être à risque pour la maladie considérée. De nombreux modèles ont été proposés pour étudier des durées de vie en présence d'une fraction immune. Récemment, une généralisation de ces modèles a été proposée par Lu et Ying (2004): elle combine un modèle de régression semi-paramétrique de transformation linéaire pour le risque de survenue de l'évènement et un modèle logistique pour la fraction immune.

L'objectif de cette thèse est d'établir rigoureusement les propriétés asymptotiques d'estimateurs du maximum de vraisemblance, pour deux modèles de régression semi-paramétriques de durées: un modèle de mélange semi-paramétrique pour risques concurrents et un modèle de régression semi-paramétrique de transformation linéaire pour durées de vie avec une fraction immune. Le premier modèle correspond à la spécification proposée par Escarela et Bowater (2008). L'étude du deuxième modèle a pour but présenter une généralisation du modèle proposé par Lu et Ying (2004). Cette généralisation comprend comme cas particuliers quelques modèles de durée de vie avec fraction immune présents dans la littérature (Farewell, 1982; Kuk et Chen, 1992; Sy et Taylor, 2000; Peng et Dear, 2000). Spécifiquement, ce travail étudie les proprités asymptotiques d'estimateurs dits du maximum de vraisemblance non-paramétrique pour les deux modèles mentionnés précédemment. Dans la suite de ce travail, nous présentons ces modèles, la méthode d'estimation proposée, et établissons les propriétés asymptotiques des estimateurs construits, à l'aide d'outils de la théorie des processus empiriques.

Les deux modèles auxquels nous nous intéressons sont des modèles semi-paramétriques (dans lesquels interviennent des paramètres fonctionnels). Dans ce travail, nous utilisons et enrichissons les techniques développées par Murphy (1994,1995) et Parner (1998) pour les modèles de fragilité. Nous étudions l'existence des estimateurs proposés, et à l'aide d'outils de la théorie des processus empiriques, nous établissons leur existence, consistance, normalité asymptotique, efficacité semi-paramétrique. Nous considérons également le problème de l'estimation de la variance asymptotique des estimateurs proposés.

Cette thèse est organisée en deux parties. La première partie comprend quatre chapitres et elle est dédiée à donner une brève introduction des thèmes qui seront utiles durant le déroulement du travail. Dans le Chapitre 1, nous rappelons les définitions des outils de modélisation utilisés en analyse statistique des durées de vie, et la notion de censure. Puis nous présentons le modèle de Cox et nous rappelons les principaux résultats

sur l'estimation de ses paramètres. Des processus de comptage et outils de martingales fournissent le guide pour étudier les propriétés asymptotiques des estimateurs pour les modèles non-paramétriques et semi-paramétriques. Les définitions et notions utiles sont rappelées dans le Chapitre 2. Le Chapitre 3 donne une description des idées principales et techniques de l'inférence semi-paramétrique en insistant sur l'efficacité semi-paramétrique. Finalement, le Chapitre 4 décrit les outils de processus empiriques qui sont importants pour étudier les propriétés asymptotiques.

La deuxième partie de la thèse comprend deux chapitres dans lesquels les deux modèles mentionnés précédemment sont étudiés. Dans le Chapitre 5, nous étudions le modèle semi-paramétrique pour risques concurrents proposé par Escarela et Bowater (2008). Des conditions de régularité et l'approche du maximum de vraisemblance et l'algorithme EM du modèle sont présentés. La théorie de processus empiriques est utilisée pour démontrer la consistance et la normalité asymptotique des estimateurs du maximum vraisemblance. Nous étudions le problème de l'estimation de la variance asymptotique des estimateurs, ainsi que l'efficacité semi-paramétrique. Enfin, nous montrons deux simulations afin de comparer le modèle de mélange semi-paramétrique pour risques concurrents par rapport à un modèle paramétrique.

Dans le Chapitre 6, nous étudions la classe générale des modèles de transformation semi-paramétriques avec une fraction immune. Nous présentons le cadre des modèles avec une fraction immune, puis une brève introduction aux modèles de transformation est présentée. Ensuite, nous développons un modèle de transformation semi-paramétrique linéaire avec covariables dépendant du temps et une fraction immune, nous donnons quelques notations et hypothèses du modèle qui sont utilisées dans les sections suivantes. Nous démontrons les propriétés d'identifiabilité du modèle, d'existence, consistance, normalité asymptotique et efficacité des estimateurs du maximum de vraisemblance non-parametriques. Finalement, nous présentons des estimateurs convergents de la variance asymptotique pour les paramètre euclidiens et infini - dimensionnel. Finalement, nous donnons quelques conclusions et projetons quelques travaux futurs.

List of symbols

Here are some general notation which will be used throughout the thesis:

- All the random variables are defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.
- $\mathbb{R}$ denotes the real numbers.
- $\mathbb{R}^k$ denotes the k-dimensional Euclidean space.
- $|\cdot|$ denotes the Euclidean norm on $\mathbb{R}^p$.
- $a \wedge b$ is the minimum between a and b.
- $a \vee b$ is the maximum between a and b.
- Let C a set,
  - $l^\infty(C)$ denotes the set of all uniformly bounded real functions on C.
  - $VB(C)$ denotes the space of the functions on C in $\mathbb{R}$, which are bounded and the bounded variation.
  - If f is a bounded function on C in $\mathbb{R}$, $\|f\|_\infty = \sup_{t\in C} |f(t)|$ denotes the supremum norm of f.
- Let A a matrix of dimension $n \times p$,
  - A' denotes the transpose of A.
  - If $n = p$ and A is invertible, A^{-1} denotes the inverse of A.
- If a is a vector of dimension p, $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ y $a^{\otimes 2} = aa'$.
- a.s. denote the almost sure convergence.
- $\xrightarrow{d}$ denote the convergence in distribution.
- $o_p(1)$ stochastic order symbols.

- caglad function is a function that is everywhere right-continuous and has left limits everywhere.

Part I

Preliminaries

Chapter 1

Survival Analysis

Survival analysis is the phrase use to describe the analysis of data in the form of times from a well-defined *time origin* until the occurrence of some particular event or *end point*. The responses consist of long time or life cycles of a phenomenon, such as time of recurrence, the duration of the effectiveness of an intervention, a specific learning time, etc. Thus, the survival is a measure of time to respond, failure, death, relapse or develop a particular disease or event. In medical research, the time origin will often correspond to the recruitment of an individual into an experimental study, such as a clinical trial to compare two or more treatments. This in turn may coincide with the diagnostic of a particular condition, the commencement of a treatment regimen, or the occurrence of some adverse event. If the end point is the death of a patient, the resulting data are literally survival times. However, data of a similar form can be obtained when the end-point is not fatal, such as the relief of pain, or the recurrence of symptoms or from example in behavioral studies in agricultural science, one often observes the time from when a domestic animal has received some stimulus until it responds with a given type action. The methodology can also applied to data from other application areas, such as the time taken by an individual to complete a task in a psychological experiment, the storage times of seeds held in a seed bank, or the lifetimes or industrial or electronic components.

Two features that present survival data are that generally not symmetrically distributed. Typically, a histogram constructed from the survival times of a group of similar individuals will tend to be positively skewed, that is, the histogram will have a longer tail to the right of the interval that contains the largest number of observations. As a consequence, it will be not reasonable to assume that data of this type have a normal distribution. This difficulty could be resolved by first transforming the data give a more symmetric distribution, for example by taking logarithms. However, a more satisfactory approach is to adopt an alternative distributional model for the original data. Also, the survival times are frequently *censored*. There are several categories of censorship, such

as, right censoring (which is when the observation ceases before the event is observed), left censoring (which is when the observation does not begin until after the event has occurred)and interval censoring(it means to know that survival has only been true sometime within an interval of time known). In this thesis focuses on right censoring.

The intention of this chapter is to give a brief introduction to survival analysis. In the section 1, we give the main definitions in the survival analysis. The section 2 presents the most common types of censoring found in survival analysis and explains how to build the likelihood function when there are censored observations. Finally, the section 3 presents the Cox proportional hazards model.

1.1 Survival function and hazard function

In summarizing survival data, the are two functions of central interest, namely the *survival function* and the *hazard function*. These functions are defined in this section. The actual survival time of an individual, t, can be regarded as the value of a continuous variable distribution, and T is called the *random variable* associated with the survival time. However, the variable T can be a positive discrete random variable and hence the following definitions and properties can be adjusted to the discrete case.

Now suppose that the random variable T has a probability distribution with underlaying *probability density function* $f(t)$. The *distribution function* of T is then given by

$$F_T(t) = \mathbb{P}\{T \leq t\} = \int_0^t f(u)du,$$

and represents the probability that the survival time is less or equal than some value t.

The *survival function*, $S_T(t)$, is defined to be the probability that the survival time is greater than to t, and so

$$S_T(t) = \mathbb{P}\{T > t\} = 1 - F(t).$$

The survival function can therefore be used to represent the probability that an individual survives from the time origin to some beyond t. (The survival function is a monotone, nonincreasing function of time).

The *hazard function* is defined as

$$\lambda_T(t) = \lim_{\Delta t \to 0} \left[\frac{\mathbb{P}\{t \leq T < t + \Delta t | T \geq t\}}{\Delta t} \right],$$

which may be interpreted as the instantaneous failure rate among those at risk. The function $\lambda_T(t)$ is also referred to as the *hazard rate*, the *instantaneous death rate*, the *force*

of mortality or the *(*force of risk). From equation above, $\lambda_T(t)\Delta t$ is the approximate probability that an individual dies in the interval $(t, t + \Delta t)$, conditional on that person having survived to time t.

From the definition of the hazard function, we can obtained some useful relationships between the survival and hazard functions,

- $\lambda_T(t) = \frac{f(t)}{S(t)}$,
- $\lambda_T(t) = -\frac{d}{dt}\{\log S(t)\}$.

From the above expressions, the survival function can be written in terms of hazard function as follows

$$S_T(t) = \exp\{-\Lambda_T(t)\},$$

where,

$$\Lambda_T(t) = \int_0^t \lambda_T(u)du,$$

is called the *integrated or cumulative hazard* or *cumulative risk function*. This function has an interpretation all its own: it measures the total amount of risk that has been accumulated up to time t.

The function $\lambda_T(t)$ is a force of mortality if and only if it satisfies the following properties:

1. $\lambda_T(x) \geq 0$, for all x.
2. $\int_0^\infty \lambda_T(x)dx = \infty$.

Remark: Given one of the five functions ($S_T(t)$, $\lambda_T(t)$, $\Lambda_T(t)$, $F_T(t)$ and $f_T(t)$) that describe the probability distribution of failure times, the other four are completely determined.Now, the above definitions are illustrated with two parametric models of survival analysis.

1. **The Exponential model.** Suppose that the survival times of n individuals, $t_1, ..., t_n$, are assumed to have an exponential distribution with mean $1/\gamma$, then this model has survival, density, hazard and cumulative hazard function of the form,

$$S_T(t) = \exp\{-\gamma t\}, \quad f_T = \gamma \exp\{-\gamma t\}, \quad \lambda_T(t) = \gamma, \quad \Lambda_T(t) = \gamma t,$$

where $\gamma > 0$ is a parameter. The exponential family has constant hazard function and the associated *lack of memory* property

$$\mathbb{P}\{T > a + t | T > a\} = \mathbb{P}\{T > t\} \qquad \text{for all } a > 0 \text{ and } t > 0.$$

The exponential model is widely used in the domain of reliability, in the context industrial. This model is studied in this context by Meeker and Escobar (1998) and Voivnov and Nikulin (1993). Historically, this model was one of the first models in survival analysis (see Johonson *et al.* (1994), Klein and Moeschberger (1997) and Lawless (1982).

2. **The Weibull model.** Suppose that T has a distribution Weibull with scale parameter γ and shape parameter α, i.e. $T \sim WEI(1/\gamma, \alpha)$. The density function is expressed by

$$f_T(t) = \alpha\gamma^\alpha t^{\alpha-1} \exp\{-(\gamma t)^\alpha\}, \qquad \alpha, \gamma > 0,$$

and the hazard and survival function have the form,

$$\lambda_T(t) = \alpha\gamma(\gamma t)^{\alpha-1}$$

$$S_T(t) = \exp\{-(\gamma t)^\alpha\}.$$

The two parameters allow the Weibull density to take a variety of shapes, and the hazard function is either monotone increasing, decreasing or constant according to wheter $\alpha > 1$, $\alpha < 1$, or $\alpha = 1$. The case $\alpha = 1$ gives the exponential distribution.

This model is used in the domain of reliability and survival analysis in medicine, application for this model may be found in Johonson *et al.* (1994).

Other parametric models of survival can be found in the literature such as, the extreme value model, the Gompertz-Makeham model, lognormal model, the model of exponentially pieces, among others (see Cox and Oakes, 1984, Klein and Moeschberger, 1997; Johonson *et al.*, 1994; Lawless , 1982, Meeker and Escobar, 1998 and Voivnov and Nikulin, 1993).

1.2 Censoring

1.2.1 Mechanisms with censoring

Survival analysis has particular problems, because the observations are censored lifetimes most of the time. The most common types of censoring found in survival analysis are (see Cox and Oakes, 1984):

1. **Type I censoring.** The event of interest is observed if it occurs before a predetermined fixed time instant C. In this case, C is a constant (censorship) prefixed by the investigator for all sample units.

2. **Type II censoring.** In this type of censoring, the study continues until the failure of the first k individuals, where k is some predetermined integer ($k < n$).

3. **Random censorin**The total period of observation is fixed, but subjects enter the study at diferent time points. Some individuals fail, some individual lost-to-follow-up, some individual still alive at the end of the study.

A patient who entered a study time at time t_0 dies at time t_0+t. However, t is unknown, either because the individual is still alive or because the individual has been lost follow-up. If the individual was last known to be alive at time $t_0 + C$, the time C is called a censored survival time. This censoring occurs after individual has been entered ino a study, that is, to the right of the least known survival time, and is therefore known as *right censoring*. The right-censored survival time is then less than the actual, but unknown, survival time.

Other types of censoring can be found in more detail in Bagdonavičius and Nikulin (chapter 8), Hubeer, *et al.* (chapter 3) and Lawless (chapter 1).

Remark. An important assumption that will made in the analysis of censored survival data is that the actual survival time of an individual, t, is independent of any mechanism that causes that individual's survival time to be censored at time C, where $C < t$.

1.2.2 Fitting a parametric model with right-censored data

In this section, we present the maximum likelihood method for survival analysis model with right censoring. For more details on this topic are available Bagdonavičius and Nikulin (chapter 4), Kalbfleisch and Prentice (chapter 3), Lawless (chapters 3-6), Meeker and Escobar (chapters 7, 8, 11), among others.

Let $f_T(t;\theta)$, $F_T(t;\theta)$, $S_T(t;\theta)$ and $\lambda_T(t;\theta)$, the density function, the distribution function, the survival function and the hazard function respectively, induced by the measure P_θ.

Suppose that the data are regarded as n pairs of observation, where the pair for the i-th individual is (t_i, δ_i), $i = 1, ..., n$, where $t_i = T_i \wedge C_i$, C_i is the variable that represents the time of censorship and $\delta_i = 1_{\{T_i \leq C_i\}}$. One observation whit death at t contribute a term of the form $f(t)$, to the overall likelihood. If a survival time is censored at time C

(in this case, $T > C$), then the observation contribute a term the form $\mathbb{P}(T > C) = S(C)$. The total likelihood function is therefore

$$L_n(\theta) = \prod_{i=1}^{n} \{f_T(t_i;\theta)\}^{\delta_i} \{S_T(t_i;\theta)\}^{1-\delta_i}.$$

Estimates of the unknown parameters in this likelihood function are then found by maximizing the logarithm of the likelihood function. An alternative expression for the likelihood function can be obtained by writing the expression above in the form

$$\prod_{i=1}^{n} \left\{ \frac{f_T(t_i;\theta)}{S_T(t_i;\theta)} \right\}^{\delta_i} S_T(t_i;\theta),$$

so that,

$$\prod_{i=1}^{n} \{\lambda_T(t_i)\}^{\delta_i} S_T(t_i;\theta).$$

This version of the likelihood function is particularly useful when the probability density function has a complicated form, as it often does.

1.3 The Cox proportional hazards model

Not only in modeling the relationship between survival rate and time is important, but also its possible relationship with different explanatory variables for each individual. A convenient way to do this is to express the force of mortality as a function of time and explanatory variables. The fundamental idea is the same as in any regression model.

The best-known model for including explanatory variables is the proportional hazards model of Cox (1972). In this model, the hazard function depending on a vector of explanatory variables $\mathbf{x}$ with unknown coefficients β is factored as

$$\lambda(t;\beta,\mathbf{x}) = \psi(\beta,\mathbf{x})\lambda_0(t),$$

where $\lambda_0(t)$ is the *baseline* hazard corresponding to $\psi(\cdot) = 1$. The nice thing about this model is that $\lambda_0(t)$, the baseline hazard, is given no particular parametrization and, in fact, can be left unestimated.

In this specification the effect of explanatory variables is to multiply the hazard λ_0 by a factor ψ which does not depend on duration t. A specification of ψ in general use is

$$\psi(\beta,\mathbf{x}) = \exp(\boldsymbol{\beta}'\mathbf{x}),$$

This specification is convenient because nonnegative of ψ does not impose restrictions on β and estimation and inference are straightforward. As well see, estimation of β in this model does not require specification of the baseline hazard λ_0.

With the specification of proportional risks, we have

$$\log \frac{\lambda(t;\beta,\mathbf{x}}{\lambda_0(t)} = \beta'\mathbf{x}$$

i.e., the model defines the logarithm of relative risk as a linear function of covariates. Therefore, unlike the relative risk of risk itself, does not depend on time or, put another way, is constant over time (hence the name proportional hazard model).

Using proportional hazards model the survival function is given by:

$$S(t;\mathbf{x}) = \exp\{-\Lambda_0(t)\exp(\boldsymbol{\beta}'\mathbf{x})\}$$

where $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ is the integrated baseline hazard.

1.3.1 Fitting the proportional hazards model

Fitting the Cox proportional hazards model to an observed set of survival data entails estimating the unknown coefficients vector of the explanatory variable,$\mathbf{x}$, in the linear component of the model, β. The baseline hazard function, $\lambda_0(t)$, may also need to be estimated. In turns out that these two components of the model can be estimated separately. The β vector is estimated first and these estimates are then used to construct an estimate of the baseline hazard function.

The estimation process for this model, Cox (1972) introduced the partial likelihood method to estimate β, which is based on the product of the likelihoods of all the changes.

Suppose that data are available for n individuals, among whom there are r distinct death times and $n-r$ right-censored survival times. We will for the moment assume that only one individual dies at each death time, so that there are no *ties* in the data. The r ordered death times will be denoted by $t_{(1)} < t_{(2)} ... < t_{(r)}$, so that $t_{(j)}$ is the j-th ordered death time. The set of individuals who are at risk at time $t_{(j)}$ will be denoted by $R(t_{(j)})$, so that $R(t_{(j)})$ is the group of individuals who are alive and uncensored at a time just prior to $t_{(j)}$. The quantity $R(t_{(j)})$ s called *risk set*.

Cox(1972) showed that the relevant likelihood function for the proportional hazards model is given bye

$$L(\beta) = \prod_{j=1}^{r} \frac{\exp(\beta'\mathbf{x}_j)}{\sum_{l\in R(t_j)}\exp(\beta'\mathbf{x}_l)},$$

in which $\mathbf{x}_j$ is the vector of covariates for the individual who dies at the j-th ordered death time, $t_{(j)}$. The summation in the denominator of this likelihood function is the sum of the values of $\exp(\beta'\mathbf{x})$ over all individuals who are at risk time $t_{(j)}$. Note that the product is taken over the individuals for whom death times have been recorded. Individuals for whom the survival time are censored dot not contribute to the numerator of the log-likelihood function, but they do enter into summation over the risk sets at death times that occur before a censored time. Moreover, the likelihood function depends only on the ranking of the death times, since this determines the risk set at each death time. Consequently, inference about the effect of explanatory variables on the hazard function depend only on the rank order of the survival times.

Remark. Andersen and Gill (1982) showed that Cox's partial likelihood can be treated as an ordinary likelihood or as likelihood function concentrated with respect to the baseline survival distribution. The efficacy loss from using partial rather than full likelihood was studied by Efron (1977). He found that estimates derived using the partial likelihood approach are typically both consistent and efficient.

Chapter 2

Probabilistic theory

2.1 Preliminaries

Event time data, where one is interested in the time to a specific event occurs, are conveniently studied by the use of certain stochastic process. The data itself may be described as a counting process, which is simply a random function of time t, $N(t)$. It is zero at time zero and constant over time except that it jumps at each point in time where an event occurs, the jumps being of size 1.

Counting process and martingale methods provide direct ways of studying the large sample properties of estimators for rather general nonparametric and semiparametric models.

Before given the definitions and properties of counting process and martingales, we need to introduce some concept from general stochastic process theory.

Behind all theory to be developed is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ is a σ-field and $\mathbb{P}$ is probability measure defined on $\mathcal{F}$.

Definition 2.1.1 *A stochastic process is a family of random variables indexed by time* $\{X(t) : t \in \Gamma)$ *indexed by a set* Γ*, all defined on the same probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*.*

The mapping $t \to X(t, \omega)$, for $\omega \in \Omega$ is called a simple path or trajectory of X. The stochastic process X induce a family of increasing sub-σ-fields by

$$\mathcal{F}_t^X = \sigma\{X(s) : 0 \leq s \leq t\}$$

called the internal history of X.

Definition 2.1.2 *A history or filtration $(\mathcal{F}_t; t \geq 0)$ is a family of sub-σ-fields such that, for all $s \leq t, \mathcal{F}_s \subset \mathcal{F}_t$ which means $\mathcal{A} \in \mathcal{F}_s$ implies $\mathcal{A} \in \mathcal{F}_t$.*

Sometimes filtrations are combined and for two filtrations $(\mathcal{F}_t^1)$ and $(\mathcal{F}_t^2)$, so $\mathcal{F}_t^1 \vee \mathcal{F}_t^2$ denote the smallest filtration that contains both $\mathcal{F}_t^1$ and $\mathcal{F}_t^2$.

Definition 2.1.3 *A stochastic process X is adapted to a filtration $(\mathcal{F}_t)$ if, for every $t \geq 0$. $X(t)$ is $\mathcal{F}_t$-measurable, and in this case $\mathcal{F}_t^X \subset \mathcal{F}_t$.*

A nonnegative random variable T is called a stopping time with respect to $(\mathcal{F}_t)$ if $(T \leq t) \in \mathcal{F}_t$ for all $t \geq 0$. For a stochastic process X and a stopping time T, the stopped process X^T is defined by $X(t) = X(t \wedge T)$.

2.2 Martingales

Martingales play an important role in the statistical applications.

Definition 2.2.1 *A martingale with respect to a filtration $(\mathcal{F}_t)$ is a right-continuous stochastic process M with left-hand limits that, in addition to some technical conditions,*

- *M is adapted to $\mathcal{F}_t$*
- *$E|M(t)| < \infty$ for all t*
- *possesses the key martingale property*

$$E(M(t)|\mathcal{F}_s) = M(s) \qquad \text{for all } s \leq t, \tag{2.1}$$

thus starting that the mean of $M(t)$ given information up to time s is $M(s)$.

The equation (2.1) is equivalent to

$$E(dM(t)|\mathcal{F}_{t-}) = 0 \qquad \text{for all } t \geq 0. \tag{2.2}$$

where $\mathcal{F}_{t-}$ is the smallest σ-algebra containing all $\mathcal{F}_s$, $s < t$ and $dM(t) = M((t+dt)-) - M(t-)$. A martingale thus has zero-mean increments given the past, and without conditioning. The second condition of the definition above is referred to as M being *integrable*.

If M satisfies

$$E(M(t)|\mathcal{F}_s) \geq M(s) \qquad \text{for all } s \leq t,$$

instead of 2.1, then M is a submartingale. The similar form if M satisfies

$$E(M(t)|\mathcal{F}_s) \leq M(s) \qquad \text{for all } s \leq t,$$

then M is a supermartingale.

A martingale is called *square integrable* if $\sup_t E(M(t)^2) < \infty$. A *local martingale* M is a process such that there exist a localizing sequence of stopping times (T_n) such that for each n, M^{T_n} is a martingale. If, in addition M^{T_n} is a square integrable martingale, then M is said to be a *local square integrable martingale*.

To be able to formulate the Doob-Meyer decomposition we need to introduce the notation of a predictable process.

Definition 2.2.2 *A process X is predictable if and only if $X(T)$ is $\mathcal{F}_T$-measurable for all stopping times T.*

Let X be a cadlag adapted process. Then A is said to be the *compensator* of X if A is a predictable, cadlag and finite variation process such that $X - A$ is a local zero-mean martingale. If a compensator exists, it is unique.

2.3 Counting process

An alternative approach to developing inference procedures for censored data is by using counting process methodology. This approach was first developed by Aalen (1975) who combined elements of stochastic integration, continuous time martingale theory and counting process theory into a methodology which quite easily allows for development of inference techniques for survival quantities based on censored data. For more rigorous survey of this area see Andersen *et al.* (1993) and Fleming and Harrington (1991).

Definition 2.3.1 *A counting process $\{N(t)\}$ is stochastic process that is adapted to a filtration $(\mathcal{F}_t)$, cadlag, with $N(0) = 0$ and $N(t) < \infty$ a.s. and whose paths are piecewise constant with jumps of size 1.*

Given a right-censored sample, the process, $N_i(t) = I\{T_i \leq t, \delta_i = 1\}$, which are zero until individual i dies and then jumps with jumps of size 1, are counting process. The processes $N(t) = \sum_{i=1}^{n} N_i(t) = \sum_{t_i \leq t} \delta_i$ is also a counting process. This process simply counts the number of deaths in the sample at or prior to time t. The counting process gives us information about when events occur.

In the case of right-censored data, the filtration at time t, $(\mathcal{F}_t)$, consists of knowledge of the pairs (T_i, δ_i) provided $T_i \leq t$ and the knowledge that $T_i > t$ for those individuals still under study at time t. We shall denote the filtration at an instant just prior to time t by $(\mathcal{F}_{t-})$. The filtration $\{\mathcal{F}_t, t \geq 0\}$ for a given problem depends on the observer of the counting process.

For right-censored data, if death times T_i and censoring times C_i are independent, then, the change of an event at time t, given the history just prior to t, is given by

$$\begin{aligned} &\mathbb{P}[t \leq T_i \leq t + dt, \delta_i = 1 | \mathcal{F}_{t-}] \qquad (2.3) \\ &= \begin{cases} \mathbb{P}[t \leq T_i \leq t + dt, C_i > t + dt_i | T_i \geq t, C_i \geq t] = \lambda(t)dt & \text{si } T_i \geq t \\ 0 & \text{si } T_i < t \end{cases} \end{aligned}$$

For a given counting process, we define $dN(t)$ to be the change in the process $N(t)$ over a short time interval $[t, t + dt)$. $dN(t)$ is one if a death occurred at t or 0, otherwise. If we define the process $Y(t)$ as the number of individuals with a study time $T_i \geq t$, then

$$E(dN(t)|\mathcal{F}_{t-}) = Y(t)\lambda(t)dt.$$

The process $h(t) = Y(t)\lambda(t)$ is called the *intensity process* of the counting process. $h(t)$ is itself a stochastic process that depends on the information contained in the history process, $\mathcal{F}_t$ through $Y(t)$. The stochastic process $Y(t)$ is the process which provides us with the number of individuals at risk at a given time.

For the absolute continuous case, we define the process

$$H(t) = \int_0^t h(s)ds,$$

this process, called the *cumulative intensity process*, has the property that

$$E(N(t)|\mathcal{F}_{t-}) = E(H(t)|\mathcal{F}_{t-}) = H(t).$$

The last equality follows because, once we know the history just prior to t, the value of $Y(t)$ is fixed and, hence, $H(t)$ is nonrandom.

Now, we present the Doob-Meyer decomposition theorem. This theorem states that for any right-coninuous nonnegative submartingale N there is a unique increasing-right continuous predictable process H such that $H(0) = 0$ and $M + H(t)$ is a martingale.

Theorem 2.3.1 (Doob-Meyer Decomposition.) *(Fleming and Harrington, 1991 p. 37). Let N be a right-continuous nonnegative submartingale with respect to a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F} : t \geq 0\}, \mathbb{P})$. Then there exists a right-continuous martingale M and an increasing right-continuous predictable process H such that $E(H(t)) < \infty$ and*

$$N(t) = M(t) + H(t) \qquad a.s.$$

for any $t \geq 0$. If $H(0) = 0$ a.s., and if $N = M' + H'$ is another such decomposition with $H'(0) = 0$, then for any $t \geq 0$,

$$\mathbb{P}\{M'(t) \neq M(t)\} = 0 = \mathbb{P}\{H'(t) \neq H(t)\}.$$

If in addition N is bounded, then M is uniformly integrable and H is integrable.

Chapter 3

Semiparametric models

In this chapter we present an overview of the main ideas and techniques for the semiparametric models emphasizing aspects of the asymptotic efficiency. For more details of the theory presented in this chapter the reader can see the works of Bickel *et. al.* (1993), van der Vaart (1998) and Tsiatis (2006).

A *semiparametric model* $\mathcal{P} = \{P_\theta\}$ is a statistical model which has euclidean parameters and one or more infinite-dimensional parameters (for example, a real-valued function). Such models can be parametrized as $\theta = (\eta, \Lambda(\cdot)) \to P_{\eta,\Lambda(\cdot)}$, where η corresponds to the euclidean parameter and $\Lambda(\cdot)$ runs through an infinite-dimensional set and is considered a nuisance parameter. We denote by $\hat{\theta}_n$ the estimator of θ.

For example, we can considerer the Cox proportional hazard model. This model specifies the following distribution function for T (the failure time) as

$$F(t) = 1 - \exp\left(-\int_0^t \lambda_0(u)e^{\beta'\mathbf{x}}\mathrm{d}u\right),$$

where $\mathbf{x}$ is the q-dimensional covariate vector. In this case, $\theta = (\beta, \lambda_0(u))$. Note that β the unknown q-dimensional parameter vector of interest, and $\lambda_0(u)$ is a unknown non-negative function, which is an infinite-dimensional parameter.

Note that the estimation under the semiparametric model $\mathcal{P}$ is more difficult than the estimation under any parametric submodel. Then, we considerer the estimation under the parametric submodels. We say that $\mathcal{P}_0$ is a parametric submodel of the semiparametric model $\mathcal{P}$ if it satisfies the following conditions:

- $\mathcal{P}_0 \subset \mathcal{P}$.

- The parametric submodel $\mathcal{P}_0$ contains the true value.

For every parametric submodel we can calculate the Fisher information for estimating θ. It follows that the information for $\hat{\theta}_n$ is not larger than the infimum of the information over all parametric submodels, and thus $\hat{\theta}_n$ is *semiparametric efficient.*

Usually in the semiparametric models is sufficient to consider one-dimensional parametric submodels of the form $\{P_{\eta+ta,\Lambda_t}\}$ which are differentiable in quadratic mean. Then, we can obtained that

$$\frac{\partial}{\partial t} P_{\eta+ta,\Lambda_t}\bigg|_{t=0} = a' S_{\eta,\Lambda} + S_{\Lambda}$$

where $S_{\eta,\Lambda}$ is the score function for η and S_{Λ} is the score function for Λ and is considered as the tangent $\dot{\mathcal{P}}$ set for Λ.

Let $\prod_{\eta,\Lambda}$ the orthogonal projection onto $\dot{\mathcal{P}}$ in L_2. Then the efficient score function for η is

$$\tilde{S}_{\eta,\Lambda} = S_{\eta,\Lambda} - \prod_{\eta,\Lambda} S_{\eta,\Lambda}$$

and its covariance matrix is defined by

$$\tilde{I}_{\eta,\Lambda} = P_{\eta,\Lambda} \tilde{S}_{\eta,\Lambda} \tilde{S}'_{\eta,\Lambda}.$$

We denoted by $\psi_{\eta,\Lambda} = \tilde{I}^{-1}_{\eta,\Lambda} \tilde{S}_{\eta,\Lambda}$ *efficient function.* Then for a random sample $X_1, ..., X_n$ i.i.d. from a distribution that is known to belong to a set $\mathcal{P}$. We say that an estimator is *asymptotically efficient* for η if $\hat{\eta}_n$ satisfies

$$\begin{aligned} \sqrt{n}(\hat{\eta}_n - \eta) &= \sqrt{n}\mathbb{P}_n \psi_{\eta,\Lambda} + o_P(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{\eta,\Lambda}(X_i) + o_P(1). \end{aligned} \tag{3.1}$$

Chapter 4

Empirical process

In this chapter, we briefly introduce some basic results from the theory of empirical processes. These research techniques are useful for studying large sample properties of statistical estimates from models as well as for developing new and improved approaches to statistical inference. Most of the topics covered in this chapter will be developed more fully in later sections of the dissertation. For more details about the theory, see Huber and Lecoutre (1989), van der Vaart (1998) and van der Vaart and Wellner (1996).

4.1 Introduction to empirical process

An *empirical process* is a stochastic process based on a random sample $X_1, \ldots X_n$ of independent draws from a probability measure P on arbitrary probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Definition 4.1.1 *For each $\omega \in \Omega$ and each integer $n \geq 1$, the empirical measure $\mathbb{P}_n$ is defined as*

$$\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i(\omega)},$$

where δ_x is the Dirac measure at point x, that is

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases},$$

for any measurable set A.

Let $\mathbb{P}_n(A)$ the empirical measure on a borel set A. The empirical measure of a sample of random elements $X_1, ..., X_n$ in a measurable space $(\Omega, \mathcal{A})$ is the discrete measure given by

$$\mathbb{P}_n(A) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}(A) = \frac{card\{X_i \in A : i = 1, \ldots, n\}}{n}.$$

We write $\mathbb{F}_n$ as the distribution function (random) defined by

$$\mathbb{F}_n(x) = \mathbb{F}_n(x)(\omega) = \mathbb{P}_n(]-\infty, x])(\omega),$$

for all $x \in \mathbb{R}$ and $\omega \in \Omega$. Note that

$$\mathbb{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq x\}, \quad x \in \mathbb{R}.$$

The function $\mathbb{F}_n$ is called the empirical distribution function and the corresponding empirical process is $\sqrt{n}(\mathbb{F}_n - F)$. The empirical distribution function is the natural estimator for the underlying F if this is completely unknown.

For each $x \in \mathbb{R}$, it follows from the Law of Large Numbers that

$$\mathbb{F}_n(x) \overset{a.s.}{\to} F(x).$$

Moreover, the Central Limit Theorem guarantees that

$$\sqrt{n}(\mathbb{F}_n(x) - F(x)) \overset{d}{\to} N\left(0, F(x)(1 - F(x))\right).$$

Two of the basic results concerning to terms $\mathbb{F}_n$ and $\sqrt{n}(\mathbb{F}_n(x) - F(x))$ are the *Glivenko-Cantelli theorem* and the *Donsker theorem*. The first theorem extends the Law of Large Numbers and it gives uniform convergence.

Theorem 4.1.1 (Glivenko-Cantelli) *Let $X_1, X_2, ...$ be i.i.d. real-valued random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with distribution function F. Then*

$$\|\mathbb{F}_n - F\|_\infty = \sup_{x \in \mathbb{R}} |\mathbb{F}_n(x) - F(x)| \overset{a.s.}{\to} 0.$$

The Donsker's theorem give the converge in distribution of the empirical process $\sqrt{n}(\mathbb{F}_n - F)$ to a Gaussian process.

Theorem 4.1.2 (Donsker) *Let $X_1, X_2, ...$ be i.i.d. real-valued random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with distribution function F. Then, the sequence of empirical process $\sqrt{n}(\mathbb{F}_n - F)$ converges in distribution on the space $D[-\infty, \infty]$ of the cadlag functions (continuous on the right, limit on the left) to a Gaussian process $\mathbb{G}_F$ with mean zero and covariance given by,*

$$F(s \wedge t) - F(s)F(t).$$

The process $\mathbb{G}_F$ is known as *F-Brownian bridge*. There are the generalizations of Theorems 4.1.1 and 4.1.2 for a set of measurable functions, which allow to obtain the *Glivenko-Cantelli class* and the *Donsker class* of functions. Both classes intervene in the study of processes indexed by a set of functions. We present notations of these classes that will be useful in the following chapters (see van der Vaart and Wellner, 1996).

Let $X_1, \ldots, X_n$ be a random sample from a probability distribution P on a mesurable space (χ, ε). For a real-valued measurable function defined on (χ, ε), we write

$$\mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^{n} f(X_i).$$

Then $\{\mathbb{P}_n(f) : f \in \mathcal{F}\}$ is the emprical measure indexed by $\mathcal{F}$, while $\{\mathbb{G}_n(f) : f \in \mathcal{F}\}$ is the empirical process indexed by $\mathcal{F}$, where

$$\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n - P)f = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i) - Pf\right).$$

If $f \in L_1(P)$, so $Pf = \int f dP < \infty$, then it follows from the Law of Large Numbers that

$$\mathbb{P}_n(f) \xrightarrow{a.s.} P(f). \tag{4.1}$$

Suppose that $\mathcal{F}$ is a collection of real-valued functions $f : \chi \to \mathbb{R}$. If the convergence in the equation (4.1) holds uniformly over $f \in \mathcal{F}$,

$$\|\mathbb{P}_n f - Pf\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |\mathbb{P}_n f - Pf| \xrightarrow{a.s.} 0,$$

then we call $\mathcal{F}$ a *Glivenko-Cantelli class*.

If $f \in L_2(P)$, so $Pf^2 = \int f^2 dP < \infty$, then

$$\sqrt{n}(\mathbb{P}_n - P)(f) \xrightarrow{d} N\left(0, P(f - Pf)^2)\right), \tag{4.2}$$

by the Central Limit Theorem. Suppose that $\mathcal{F}$ is a collection of real-valued functions $f : \chi \to \mathbb{R}$. If the convergence in the equation (4.2) holds uniformly over $f \in \mathcal{F}$, then

$$\sqrt{n}(\mathbb{P}_n - P)(f) \Rightarrow G(f) \quad \text{in } \ell^\infty(\mathcal{F})$$

where G is a mean-zero Brownian bridge process with covariance function

$$cov(G(f), G(g)) = Pfg - PfPg.$$

Then,we say that $\mathcal{F}$ is a *Donsker class*. Here

$$\ell^\infty(\mathcal{F}) = \left\{ x : \mathcal{F} \to \mathbb{R} | \|x\|_\mathcal{F} = \sup_{f \in \mathcal{F}} |x(f)| < \infty \right\}.$$

4.2 Examples of Donsker classes

There are a number of methods which can be used to determine a Donsker class, van der Vaart (1998) and van der Vaart and Wellner (1996) give the following examples:

Example 1. If $\mathcal{F}$ is equal to the collection of all indicator functions of the form $f_t = 1\{(-\infty, t]\}$ with $t \in \mathbb{R}$. Then $\mathcal{F}$ is a Donsker class (see example 19.6 in van der Vaart, 1998).

Example 2. The set of uniformly bounded functions and uniformly bounded variation is Donsker (see example 19.11 in van der Vaart, 1998).

Example 3. The class of functions whose derivatives up to order k exist and are uniformly bounded by contants M_k is Donsker (see example 19.9 in van der Vaart, 1998).

However, we can build Donsker classes from well-known Donsker classes. For example,

Example 4. If $\mathcal{F}$ is a Donsker classe and $\sup_{f \in \mathcal{F}} |Pf| < \infty$ and $\phi : \mathbb{R} \to \mathbb{R}$ is a lipschitz function then the class of all functions of the form $\phi(f)$ is a Donsker class, if f ranges over Donsker class $\mathcal{F}$ with integrable envelope functions (see example 2.10.6 in van der Vaart and Wellner, 1996).

Example 5. If $\mathcal{F}$ and $\mathcal{G}$ are Donsker classes and $sup_{\mathcal{F} \cup \mathcal{G}} |Pf| < \infty$, the following are also Donsker: the pairwise infima $\mathcal{F} \wedge \mathcal{G}$, the pairwise suprema $\mathcal{F} \vee \mathcal{G}$, and pairwise sums $\mathcal{F} + \mathcal{G}$ (see example 2.10.7 in van der Vaart and Wellner, 1996).

Example 6. If $\mathcal{F}$ and $\mathcal{G}$ are uniformly bounded Donsker classes, then $\{fg : f \in \mathcal{F}, g \in \mathcal{G}\}$ is a Donsker class (see example 19.20 in van der Vaart, 1998).

Example 7. If $\mathcal{F}$ is Donsker with $\sup_{f\in\mathcal{F}}|Pf| < \infty$ and f is uniformly bounded away from zero for every $f \in \mathcal{F}$, then $1/\mathcal{F} = \{1/f : f \in \mathcal{F}\}$ is Donsker (see example 2.10.9 in van der Vaart and Wellner, 1996).

Example 8. If Z is a caglad process on $[0, t]$, $t \in \mathbb{R}$ which is uniformly bounded in variation, then $Z(\cdot)$ is Donsker (see Lemma 2 in Parner, 1998).

Example 9. If $\mathcal{F}$ is Donsker, then it is also Glivenko-Cantelli.

Part II

Semiparametric Mixture Model for Competing Risks and Semiparametric Transformation Cure Model

Chapter 5

Semiparametric Mixture Model for Competing Risks

Introduction

The modeling and fit of competing risks data is one of the most prominent areas of research in survival analysis from disciplines such as epidemiology, finance, criminology and engineering. For a discussion of several inference problems in competing risks see, for example, Holt (1978), Whitmore (1986), Spivey and Gross (1991), Gaynor *et al.* (1993), Klein and Moeschberger (1997) and Klein and Bajorunaite (2004) and the references therein. In the competing risks setting, the inference centres on analyzing durations of time of certain events, beginning with a well established origin in time and ending with the occurrence of the event which is classified in several cases. For example, when a company hires life insurance for their employees, the amount of the compensation varies according to the cause of death; generally a death related to the working conditions tends to be larger than that from another type; so it is important to estimate the probabilities associated with each cause in the presence of auxiliary variables, such as age and gender, to allow for the calculation of the appropriate premium. In many practical situations the data are contain concomitant information which is thought it influences the occurrence of the events.

There are different reasons for competing risks data not to be studied using standard statistical methods, mainly because the occurrence of a cause does not allow to observe the occurrence of another and, as for an other survival data, it is common to find that for some experimental units in the sample the failure times are not observed at the end of follow-up. Several regression models for analysis competing risks have been proposed (among them are: Holt JD, 1978; Prentice RL *et al.*, 1978; Spivey LB *et al.*, 1991; Gaynor *et al.*,

1993 and Fine, 1999); however, these formulations can be criticized for its assumptions unjustified and complicated interpretation.

A problem with the modeling of competing risks is that there is not always a cause-specific hazard function that in an unique way identifyes the joint survival function. To solve this problem of identifiability (see Tsiatis, 1975), several models assume that events are mutually independent; however, this assumption can be questionable. Carrière (1995) proposed to use a copula function to model the joint survival function which allows a structure of dependency between the risks and therefore the competing risks model is identifiable. A problem to this formulation given by Carrière is that there are few multivariate copulas, which restricts the model to two risks, moreover, there is no methodology to determine what copula and what marginal are appropriate to fit the data.

There are two extensions of Cox's proportional hazards model for analysing cause-specific survival data that have been generally regarded as being particularly important. The first, described by Kalbfleisch and Prentice (1980), consists of fitting the standard proportional hazards model separately for each type of failure in turn, treating other failure types as censored data. Since this method does not include all types of failures simultaneously in the inference process, the interpretation of parameters estimates is complicated. Furthermore, the model implies an infinite-dimensional specification for the cause-specific hazard functions. Therefore, if for instance, interest lies in estimating the cause-specific survival probabilities, the resulting estimated hazards will generally have very wide confidence bands (see for example Cheng *et al.*, 1998).

An alternative approach was presented by Larson and Dinse (1985). Their model incorporates the different failure types by splitting the population into groups of individuals who eventually fail from each cause with probabilities being attributed to the membership of each group. In addition, they advocate that the effects of covariates on each group are investigate through a parametric proportional hazards regression. Although complex, the use of this mixture model can be considered as being an attractive statistical approach as it is a model that is fully specified and easy to interpret. Larson and Dinse (1985) developed a maximum likelihood estimation procedure for their model, and Choi and Zhou (2002) and Maller and Zhou (2002) investigated large-sample properties of the resulting estimators, which include existence, consistency, and asymptotic normality.

An attempt to generalize Larson and Dinse's model to a semiparametric context was carried out by Kuk (1992). He proposed a mixture model analogous to the standard Cox proportional hazards model where the baseline hazard functions are eliminated as nuisance parameters. The inferences depend on a Monte Carlo approximation of the likelihood function involved, which has a number of drawbacks since it is computationally expensive to carry this out and the corresponding standard errors are sensitive to the sampling plan used. Other semiparametric generalizations of Larson and Dinse's model have recently been proposed and investigated. For example, Ng vesand McLachlan (2003)

and Escarela and Bowater (2008) present a class of model specifications that includes Kuk's mixture model. They consider a semiparametric mixture model with covariates, where the conditional marginals of each cause have a multinomial logistic form and the conditional distribution of the time given the covariates and the failure cause is specified through a proportional hazards model (Cox, 1972). Ng and McLachlan (2003) and Escarela and Bowater (2008) propose and implement EM-type algorithms in this class of models. Naskar *et al.* (2005) consider a similar model in the case of clustered failure time data, and develop a Monte Carlo EM algorithm.

The aforementioned papers focus on the computational aspects of the estimation in semiparametric mixture models for competing risks data. To the best of our knowledge, only a few papers have contributed to the large-sample properties of the estimators in these contributions models. Dupuy and Escarela (2007) outline a consistency proof for the maximum likelihood-based estimators proposed by Escarela and Bowater (2008). Lu and Peng (2008) construct martingale-based estimating equations for the parameters of a semiparametric version of Larson and Dinse's model, and establish the consistency and asymptotic normality of the resulting estimators.

In this chapter, we focus on the properties of the maximum likelihood-based estimators in the semiparametric generalization of Larson and Dinse's model developed by Escarela and Bowater (2008). Specifically, we provide a rigorous large-sample treatment of the resulting estimators. By following the approach and techniques developed by Murphy (1994, 1995) and Parner (1998) for the frailty model (and thereafter extended to various other settings by Fang et al. (2005), Dupuy et al. (2006), Kosorok and Song (2007), Lu (2008), among others), we prove the consistency and asymptotic normality of the estimators in Escarela and Bowater (2008). We also show that the proposed estimator for the regression parameter of interest, which is the regression parameter in the conditional distribution of the failure time given the failure cause and covariates, is semiparametric efficient. Consistent variance estimators are also obtained.

Introduction

La modélisation de durées de vie dans un contexte de risques concurrents intervient dans de nombreuses disciplines: épidémiologie, fiabilité, finance, criminologie,... (voir par exemple, Holt (1978), Whitmore (1986), Spivey et Gross (1991), Gaynor *et al.* (1993), Klein et Moeschberger (1997) et Klein et Bajorunaite (2004)).

Le problème consiste à analyser un échantillon de durées (jusqu'au décès, par exemple), lorsque le décès peut être dû à plusieurs causes mutuellement exclusives. Par exemple, considérons une entreprise qui passe un contrat d'assurance-vie pour ses employés. Les montants des indemnisations vont varier suivant la cause de la mort du travailleur. Une mort relative aux conditions de travail tendra à engendrer un coût plus important. Il est donc important d'estimer les probabilités associées à chacune des causes de décès, y compris en présence de variables auxiliaires importantes, comme l'âge et le sexe.

Il existe différentes raisons par lesquelles les données de risques concurrents ne peuvent pas être étudiées par les méthodes statistiques standards. Une raison est que la survenue d'une cause ne permet pas d'observer la survenue des autres. De plus, il est fréquent que la durée d'intérêt soit censurée. Dans ce cas, la cause de l'évènement à venir sera inconnue. Plusieurs modèles de régression pour l'analyse de risques concurrents ont été proposés (entre ceux-ci se trouvent les formulations faites par Holt,1978; Prentice *et al.*, 1978; Spivey et al., 1991; Gaynor *et al.*, 1993; Carrire, 1995 et Fine, 1999). Toutefois, ces formulations peuvent être critiquées, à cause de leurs hypothèses contraignantes et de l'interprétation compliquée de leurs résultats.

Dans la littérature récente, plusieurs extensions du modèle de Cox aux risques concurrents ont été proposées. La proposition faite par Kalbfleish et Prentice (1973) consiste à adopter le modèle à risques proportionnels pour chaque type de risque, où le traitement des autres types de risques est pris comme censuré. Puisque cette méthode n'inclut pas tous les risques simultanément dans le processus d'inférence, le modèle peut avoir de très amples bandes de confiance pour les probabilités de survie de cause spécifique (voir par exemple Fine,1999).

Un point de vue alternatif a été présenté par Larson et Dinse (1985). Cette formulation repose sur un modèle multinomial généralisé pour les probabilités de la cause d'évènement, et le modèle de régression exponentiel par morceaux pour les fonctions conditionnelles de survie de chaque cause. Deux inconvénients de ce modèle sont que: lorsqu'il est paramétrique, il est peu flexible, et choisir des formes paramétriques incorrectes pour les fonctions conditionnelles de survie implique d'obtenir des inférences erronées. Larson et Dinse (1985) ont développé la méthode du maximum de vraisemblance pour leur modèle de mélange paramétrique. Choi et Zhou (2002) et Maller et Zhou (2002) ont étudié théoriquement les propriétés asymptotiques de ces estimateurs.

Une généralisation semi-paramétrique du modèle de Larson et Dinse (1985) a été proposée par Kuk (1992), dont le modèle consiste à spécifier les fonctions conditionnelles de survie par le modèle semi-paramétrique de Cox. Récemment, d'autres généralisations semi-paramétriques du modèle de Larson et Dinse (1985) ont été proposées. Par exemple, Ng et McLachlan (2003) et Escarela et Bowater (2008) présentent une classe de modèle qui inclut le modèle de Kuk. Ils considèrent un modèle de mélange semi-paramétrique avec covariables, où la distribution conditionnelle de la cause d'évènement a une forme logistique, et la distribution conditionnelle du temps de décès sachant les covariables et la cause de décès est spécifiée par un modèle de risques proportionnels (Cox, 1972). Naskar *et al.*. (2005) considèrent un modèle similaire dans le cas où les individus de l'échantillon se répartissent en clusters, et développent un algorithme Monte-Carlo EM.

Les articles mentionnés ci-dessus se concentrent seulement sur les aspects algorithmiques de l'estimation du modèle de mélange semi-paramétrique pour la modélisation de risques concurrents. Seulement quelques articles ont contribué aux propriétés asymptotiques des estimateurs proposés dans ces contributions. Dupuy et Escarela (2007) décrivent une preuve de la consistance pour les estimateurs du maximum de vraisemblance proposés par Escarela et Bowater (2008). Lu et Peng (2008) construisent des équations d'estimation en utilisant des martingales, et établissent la consistance et la normalité asymptotique de leur estimateurs.

Dans ce chapitre, nous nous concentrerons sur les propriétés asymptotiques des estimateurs du maximum de vraisemblance de la généralisation semi-paramétrique du modèle de Larson et Dinse (1985) développée par Escarela et Bowater (2008). Spécifiquement, nous fournissons un traitement rigoureux des propriétés asymptotiques des estimateurs résultants. Nous utiliserons et adapterons les techniques développées par Murphy (1994, 1995) et Parner (1998) pour le modèle de fragilité. Nous démontrons l'existence, la consistance, la normalité asymptotique des estimateurs proposés par Escarela et Bowater (2008). Nous montrons aussi que l'estimateur proposé pour le paramètre de régression d'intérêt est efficace au sens semi-paramétrique. Finalement un estimateur consistant de la variance asymptotique des estimateurs proposés est obtenu.

5.1 Theoretical framework of competing risks model

Consider the following theoretical framework to develop an appropriate methodology to analyze competing risks (see Eldant-Johnson and Johnson, 1980):

- Each death is due to a single cause.
- Each individual in a given population is liable to die from any of the causes operating in this population.

Consider a population in which there are J causes of death. In view of the first assumption, that each death is due to a single cause, we cannot observe $(T_1, ..., T_J)$ jointly. Instead, we observed time at death $T = \min(T_1, ..., T_J)$. Let $T_1, \dots T_J$ denote the hypothetical (potential) times due to die and define their joint survival distribution function

$$S(t_1, ..., t_J) = \mathbb{P}\{T_1 > t_1, ..., T_J > t_j\}.$$

We assume that $S(t_1, ..., t_J)$ is absolutely continuous. The force of mortality of cause-specific mortality is expressed as follows,

$$\lambda_j^*(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}\left\{t < T_j < t + \Delta t | T_j \geq t\right\}}{\Delta t},$$

and is interpreted as the instantaneous probability of occurrence of an event of type j in the time t given that the individual has survived all causes. In terms of $S(t_1, ..., t_J)$, $\lambda_j(t)$, is expressed as:

$$\lambda_j^*(t) = - \left. \frac{\partial \log S(t_1, ..., t_J)}{\partial t_j} \right|_{t_i = t}.$$

The overall survival function is

$$S_T(t) = \mathbb{P}\{T > t\} = \mathbb{P}\left(\cap_{j=1}^{J}\{T_j > t\}\right) = S(t, ..., t). \tag{5.1}$$

Define the random variable H that identifies the cause of failure as

$$H = \sum_{i=1}^{J} jI(T = T_J).$$

If we observe the pair of values (T, H), then the time at death and the cause of death are identified.

The conditional probability of death from cause j in an interval $(t, t + \Delta t)$, given alive at age t, and in the presence of all other causes acting simultaneously in a population, is approximately, $\lambda_j^*(t)dt$. The unconditional probability of death from cause j in $(t, t + \Delta t)$

is then $S_T(t)\lambda_j^*(t)dt$. Hence, the crude (in the presence of all causes) probability of time at death for cause j is

$$F_j^*(t) = \mathbb{P}\{T \leq t, H = j\} = \int_0^t \lambda_j^*(u) S_T(u) du,$$

for $j = 1, \ldots J$. Let p_j the expect proportion of deaths from cause j. We have

$$p_j = \mathbb{P}\{M = j\} = \int_0^\infty \lambda_j^*(u) S_T(u) du = F_j(\infty),$$

with $p_1 + ... + p_J = 1$.

5.2 The mixture model framework

In this seection, we describe the data and the semiparametric mixture regression model derived from Larson and Dinse (1985) proposed by Escarela and Bowater (2009). First, we consider that all the random variables are defined on a probability space $(\Omega, \mathcal{C}, \mathbb{P})$. Let T^0 be a random failure time of interest. As is custom in survival analysis, we suppose that T^0 may be right-censored by a positive random variable C (in the example of prostate cancer data, some individuals were lost to follow up during the course of the study, and were considered as right-censored). Let $\mathbf{Z}$ and $\mathbf{X}$ be respectively p- and q-vectors of covariates ($\mathbf{Z}$ and $\mathbf{X}$ may share some common components). Let H be the failure cause variable and $\mathcal{J} = \{1, \ldots, J\}$ be the set of possible values of H. For $j \in \mathcal{J}$, we define the indicator variable $\Gamma^j = 1\{H = j\}$. In a competing risks setting, both the failure cause H and the indicator Γ^j are observed only if the survival time is uncensored.

The mixture model approach proposed by Larson and Dinse (1985) assume that the cause of death of an individual is chosen at the outset by a stochastic mechanism from the J possible causes. Let

$$p_j = \mathbb{P}\{H = j\}$$

be the probability of ever failing from cause j. They assume that the survival time is a realisation of T, therefore the model is based on the *conditional survival distribution functions*, defined as

$$S_j(t) = \mathbb{P}\{T > t | H = j\}, \quad j \in \mathcal{J}$$

where $S_j(t)$ is a *proper survival distribution* in the sense that $S_j(0) = 1$ and $S_j(\infty) = 0$. It follows that the *cause-specific failure probabilities* also known as cumulative incidence function can be calculated by $F_j(t) = p_j[1 - S_j(t)]$. Thus, the *overall survival function*, which is defined by $S_T(t) = \mathbb{P}\{\bigcap_{j=1}^J (T_j > t)\}$, can be expressed as

$$S_T(t) = \mathbb{P}\{T > t\} = \sum_{j=1}^J p_j S_j(t).$$

In order to allow for the effect of covariates on the conditional survival function $S_j(t)$ or in other words their effect on the *hazard component* of the mixture model, it is convenient to make use of the Cox proportional hazards model. More specifically, it is convenient to assume that this hazard component is characterized by the following function:

$$S_j(t;\mathbf{Z}) = \exp\left\{-\Lambda_{j0}(t)e^{\beta_j'\mathbf{Z}}\right\}, \qquad j \in \mathcal{J},$$

where $\Lambda_{j0}(t) = \int_0^t \lambda_{j0}(u)du$ is the integrated baseline hazard function for the failure j, $\mathbf{Z}$ is a p-vector of covariates, which does not contain the intercept, and β_j denotes the p-vector of parameters for failure j. In terms of the hazard function, the assumption of there being proportional hazards implies that

$$\begin{aligned}\lambda_j(t;\mathbf{Z}) &= \lim_{h\downarrow 0}\frac{1}{h}\mathbb{P}(t < T^0 \leq t+h | T^0 > t, H = j, \mathbf{Z}) \\ &= \lambda_{j0}(t)\exp(\beta_j'\mathbf{Z}), \quad j = 1, ..., J. \end{aligned} \tag{5.2}$$

In similar way, the effects of covariates on the probabilities of eventual cause-specific death can be modeled using a *generalized logistic model* (Cox and Snell, 1989, pp. 155-157), so that the *probability model* has the following form:

$$p_j = \mathbb{P}(H = j|\mathbf{X}) = \frac{\exp(\gamma_j'\mathbf{X})}{\sum_{k=1}^J \exp(\gamma_k'\mathbf{X})}, \quad j \in \mathcal{J} \tag{5.3}$$

where $\mathbf{X}$ is a q-vector of covariates, which includes the intercept, and γ_j ($j \in \mathcal{J}$) is the corresponding q-dimensional vector of parameters for failure j. For identifiability purposes γ_J is set equal to 0.

The statistical problem is that of estimating the parameters β_j, γ_j, and the cumulative baseline hazard functions $\Lambda_j = \int \lambda_j$ from the incomplete data vectors $\mathbf{O}_i$, $i = 1, \ldots, n$. In practice, the coefficients β_j ($j \in \mathcal{J}$) are often the parameters of interest in the mixture model (5.2)-(5.3).

The semiparametric mixture model for competing risks has been employed by various authors. For example, Kuk (1992) analysed a heart transplant dataset (using a simplified version of model (5.2)-(5.3), where $\mathbf{X} = \mathbf{Z}$), while Ng and McLachlan, 2003 and Escarela and Bowater, 2008 fitted the present model to a prostate cancer dataset.

Although Larson and Dinse's original parametric model accounts for the same set of covariates in both the conditional hazard functions and the multinomial logistic regression model, Escarela and Bowater's semiparametric specification allows for different sets of covariates in each component, in order that the conditional hazard and multinomial models can each be considered in some sense as being parsimonious. Indeed, the authors found that, while the two factors they considered in the analysis of the prostate cancer data have significant effects in the generalized logistic model, neither of them seemed to be significant in the conditional hazards.

Note that the model (5.2)-(5.3) is related to semiparametric mixture models for survival data with cure fraction (see, among others, Kuk and Chen, 1992; Taylor, 1995; Sy and Taylor, 2000; Peng, 2003; who investigated the computational issues raised by the estimation in this class of models. See also Fang *et al.*, 2005 and Lu, 2008; who studied the large-sample properties of maximum likelihood estimators in the proportional hazards cure model).

5.3 Notation and model assumptions

In this section, we state some notations and model assumptions that will be used throughout the chapter. Some notations will be useful in the present study. Suppose that there is a random sample of size n. This will induce an index i $(i = 1, \ldots, n)$ on all the random variables defined above. The data consist of n independent vectors $(T_i, \Delta_i, \mathbf{Z}_i, \mathbf{X}_i, \Delta_i H_i)$ $(i = 1, \ldots, n)$, where $T_i = \min\{T_i^0, \min(C_i, \tau)\}$, $\Delta_i = 1\{T_i^0 \leq \min(C_i, \tau)\}$, and $\tau < \infty$ is a fixed constant denoting the end of the study. In this work, we will use $\mathbf{O}$ and $\mathbf{O}_i$ to abbreviate the observed data vector $(T, \Delta, \mathbf{Z}, \mathbf{X}, \Delta H)$ and its n independent replicates $(T_i, \Delta_i, \mathbf{Z}_i, \mathbf{X}_i, \Delta_i H_i)$.

We shall note $\mathbf{G} = (\gamma_1' \ldots \gamma_{J-1}')'$, $p_{\mathbf{G}}^{j,\mathbf{X}} = \mathbb{P}(H = j|\mathbf{X})$, and $p_{\mathbf{G},i}^{j,\mathbf{X}} = \mathbb{P}(H = j|\mathbf{X}_i)$. If $t \in [0, \tau]$, we denote by $N(t) = 1\{T \leq t\}\Delta$ and $Y(t) = 1\{T \geq t\}$ the failure counting and at risk processes respectively. For $j \in \mathcal{J}$, define the counting process $N^j(t) = 1\{T \leq t\}\Delta^j$, where $\Delta^j = \Delta\Gamma^j$. Note that $N^j(t)$ is equal to 1 if the failure arises form the j-th cause before time t. Corresponding quantities for the i-th subject will be denoted by N_i, N_i^j, and Δ_i^j.

To establish our results, we need the following regularity assumptions:

(C1) Conditionally on $\mathbf{Z}, H$, and $\mathbf{X}$, the censoring time C is independent of the failure time T^0. Conditionally on $\mathbf{Z}$ and X, C is independent of H.

(C2) There exists a positive constant c_0 such that $\mathbb{P}(C \geq \tau|\mathbf{Z}, \mathbf{X}) > c_0$ almost surely.

(C3) The hazard function of C given $\mathbf{Z}$ and $\mathbf{X}$, $\lambda_C(s|\mathbf{Z}, \mathbf{X})$, is uniformly bounded almost surely.

(C4) Let $\mathbf{B} = (\beta_1' \ldots \beta_J')' \in \mathbb{R}^{pJ} \equiv \mathbb{R}^P$ and $\mathbf{G} = (\gamma_1' \ldots \gamma_{J-1}')' \in \mathbb{R}^{q(J-1)} \equiv \mathbb{R}^Q$. The true values $\mathbf{B}_0$ of $\mathbf{B}$ and $\mathbf{G}_0$ of $\mathbf{G}$ lie in the interior of known compact sets $\mathcal{B} \subset \mathbb{R}^P$ and $\mathcal{G} \subset \mathbb{R}^Q$ respectively.

(C5) For every $j \in \mathcal{J}$, the true conditional cumulative baseline hazard $\Lambda_{j,0}$ is a strictly increasing function in $[0, \tau]$, with $\Lambda_{j,0}(0) = 0$ and $\Lambda_{j,0}(\tau) < \infty$. $\Lambda_{j,0}$ is continuously differentiable in $[0, \tau]$, with $\lambda_{j,0}(t) = d\Lambda_{j,0}(t)/dt$.

(C6) The covariate vectors $\mathbf{Z}$ and $\mathbf{X}$ are bounded that is, $\|\mathbf{X}\| < c_1$ and $\|\mathbf{Z}\| < c_1$ for some constant $0 < c_1 < \infty$ (where $\|\cdot\|$ denotes the Euclidean norm). The covariance matrices of $\mathbf{Z}$ and $\mathbf{X}$ are positive definite. Let

$$c_2 = \min_{\beta_j, j\in\mathcal{J}, \|\mathbf{Z}\|<c_1} \exp(\beta_j'\mathbf{Z}) \qquad c_3 = \max_{\beta_j, j\in\mathcal{J}, \|\mathbf{Z}\|<c_1} \exp(\beta_j'\mathbf{Z}).$$

Let $\mathcal{L}$ be the set of all functions verifying the conditions in C5, θ denote the parameter $(\mathbf{B}, \mathbf{G}, \Lambda_j; j \in \mathcal{J})$, $\theta_0 = (\mathbf{B}_0, \mathbf{G}_0, \Lambda_{j,0}; j \in \mathcal{J})$, and $\Theta = \mathcal{B} \times \mathcal{G} \times \mathcal{L}^{\otimes J}$ denote the parameter space. Under the true value θ_0, the expectation of random variables will be noted by P_{θ_0}.

(C7) There is a positive constant c_4 such that for every $j \in \mathcal{J}$, $P_{\theta_0}[Y(\tau)\Gamma^j] > c_4$.

(C8) There exists a positive constant c_5 such that for every $j \in \mathcal{J}$, $P_{\theta_0}[\Delta^j | T, \mathbf{Z}, \mathbf{X}] > c_5$.

(C9) The distribution of the failure cause H conditionally on $\mathbf{X}$ and $\mathbf{Z}$ does not involve the components of $\mathbf{Z}$ that are not in $\mathbf{X}$. The distributions of C, $\mathbf{Z}$, and $\mathbf{X}$ do not depend on θ.

Remark. Condition **C1** ensures that no information about θ is lost by removing terms adhering to censoring from the likelihood. Condition **C2** ensures that the follow-up is sufficiently long for identifying the cumulative baseline hazard functions $\Lambda_{j,0}$ on the interval $[0, \tau]$. Conditions **C3**-**C8** are used for the identifiability of θ_0 and the asymptotics of the proposed estimators. Condition **C7** ensures that the follow-up is sufficiently long (for every failure cause) so that we can estimate the $\Lambda_{j,0}$ on the entire interval $[0, \tau]$. Condition **C8** ensures that for each failure cause, failures can happen and their cause be observed at any time and for any value of the covariates. Condition **C9** ensures that no information about θ is lost by removing terms adhering to the marginal distributions of $\mathbf{Z}$ and $\mathbf{X}$ from the likelihood.

5.4 Nonparametric maximum likelihood estimation

In this paper, we assume that there are no tied failure times (this assumption is made for ease of presentation, but our results can be easily adapted to accomodate ties). Under models (5.2) and (5.3), and conditions **C1**-**C9**, the likelihood function for the parameter θ from the observations $\mathbf{O}_i$ $(i = 1, \ldots, n)$ is proportional to

$$\prod_{i=1}^{n} \left\{ \prod_{j\in\mathcal{J}} \left[\lambda_j(T_i) e^{\beta_j'\mathbf{Z}_i} \exp\left(-e^{\beta_j'\mathbf{Z}_i} \Lambda_j(T_i) \right) p_{\mathbf{G},i}^{j,\mathbf{X}} \right]^{\Delta_i^j} \left[\sum_{j\in\mathcal{J}} \exp\left(-e^{\beta_j'\mathbf{Z}_i} \Lambda_j(T_i) \right) p_{\mathbf{G},i}^{j,\mathbf{X}} \right]^{1-\Delta_i} \right\} \quad (5.4)$$

It would seem natural to calculate the maximum likelihood estimator (MLE) of θ_0 by maximizing the foregoing likelihood. However, the maximum of this function is infinity when the functions Λ_j $(j \in \mathcal{J})$ range within the class $\mathcal{L}$ of absolutely continuous cumulative baseline hazards. To see this, we may choose functions Λ_j $(j \in \mathcal{J})$ with fixed values at the failure times T_i, and let $d\Lambda_{j,0}(T_i)/dT_i = \lambda_j(T_i)$ go to infinity for some T_i with $\Delta_i^j = 1$.

To solve this problem, we restrict the functions Λ_j $(j \in \mathcal{J})$ to be right-continuous, and we allow each Λ_j to have jumps at the failure times T_i. Then, letting $\Lambda_j\{t\}$ denote the jump size of Λ_j at t, we maximize the function $L_n(\mathbf{B}, \mathbf{G}, \Lambda_j; j \in \mathcal{J}) =$

$$\prod_{i=1}^{n} \left\{ \prod_{j \in \mathcal{J}} \left[\Lambda_j\{T_i\} e^{\beta_j' \mathbf{Z}_i} \exp\left(-e^{\beta_j' \mathbf{Z}_i} \Lambda_j(T_i) \right) p_{\mathbf{G},i}^{j,\mathbf{X}} \right]^{\Delta_i^j} \left[\sum_{j \in \mathcal{J}} \exp\left(-e^{\beta_j' \mathbf{Z}_i} \Lambda_j(T_i) \right) p_{\mathbf{G},i}^{j,\mathbf{X}} \right]^{1-\Delta_i} \right\}$$

over the space $\Theta_n =$

$$\{(\mathbf{B}, \mathbf{G}, \Lambda_j) : \mathbf{B} \in \mathcal{B}, \mathbf{G} \in \mathcal{G}, \Lambda_j \text{ is an increasing right-continuous function on } [0, \tau], j \in \mathcal{J}\}.$$

If they exist, the resulting estimators will be referred to as nonparametric MLEs (NPMLEs), and will be noted by $\widehat{\theta}_n = (\widehat{\mathbf{B}}_n, \widehat{\mathbf{G}}_n, \widehat{\Lambda}_{j,n}; j \in \mathcal{J})$, where

$$\widehat{\mathbf{B}}_n = (\widehat{\beta}_{1,n}' \ldots \widehat{\beta}_{J,n}')' \quad \text{and} \quad \widehat{\mathbf{G}}_n = (\widehat{\gamma}_{1,n}' \ldots \widehat{\gamma}_{J-1,n}')'.$$

In our setting, existence of the NPMLEs is ensured by the following result:

Proposition 5.4.1 *Under conditions **C1**-**C9**, the maximizer $\widehat{\theta}_n$ of L_n over Θ_n exists and is achieved.*

Proof: We first identify the form of a possible maximizer $\widehat{\Lambda}_{j,n}$ of L_n in the space Θ_n.

Let $j \in \mathcal{J}$. Define $\mathcal{S}_n^j = \{i \in \{1, \ldots, n\} | \Delta_i^j = 1\}$ as the set of sample individuals who are observed to fail from the j-th cause. For every $j \in \mathcal{J}$ and any function Λ_j in Θ_n, we can construct an increasing step function Λ_j^* with jumps only at the failure times in $\{T_i, i \in \mathcal{S}_n^j\}$, and satisfying $\Lambda_j^*(T_i) = \Lambda_j(T_i)$. Clearly, at each of these failure times, $\Lambda_j^*\{T_i\} \geq \Lambda_j\{T_i\}$ which implies that $L_n(\mathbf{B}, \mathbf{G}, \Lambda_j; j \in \mathcal{J}) \leq L_n(\mathbf{B}, \mathbf{G}, \Lambda_j^*; j \in \mathcal{J})$. Therefore, the maximizer $\widehat{\Lambda}_{j,n}$ (if it exists) must be a step function with positive jumps at the failure times T_i such that $\Delta_i^j = 1$. This restricts the maximization problem of L_n to the following subspace of Θ_n:

$$\left\{ \left(\mathbf{B}, \mathbf{G}, \Lambda_j\{t_k^j\}\right) : \mathbf{B} \in \mathcal{B}, \mathbf{G} \in \mathcal{G}, \Lambda_j\{t_k^j\} \in [0, \infty), k = 1, \ldots, |\mathcal{S}_n^j|, j \in \mathcal{J} \right\}, \tag{5.5}$$

where for every $j \in \mathcal{J}$, $|\mathcal{S}_n^j|$ denotes the cardinality of $\mathcal{S}_n^j$ and $t_1^j < \ldots < t_{|\mathcal{S}_n^j|}^j$ are the ordered failure times in the set $\{T_i, i \in \mathcal{S}_n^j\}$. That is, we maximize the function

$$L_n(\mathbf{B}, \mathbf{G}, (\Lambda_j\{t_k^j\})_{j,k}) =$$

$$\prod_{i=1}^{n} \left\{ \prod_{j \in \mathcal{J}} \left[\Lambda_j\{T_i\} e^{\beta_j' \mathbf{Z}_i} \exp\left(-e^{\beta_j' \mathbf{Z}_i} \sum_{k=1}^{|\mathcal{S}_n^j|} \Lambda_j\{t_k^j\} 1\{t_k^j \leq T_i\} \right) p_{\mathbf{G},i}^{j,\mathbf{X}} \right]^{\Delta_i^j} \right.$$
$$\left. \times \left[\sum_{j \in \mathcal{J}} \exp\left(-e^{\beta_j' \mathbf{Z}_i} \sum_{k=1}^{|\mathcal{S}_n^j|} \Lambda_j\{t_k^j\} 1\{t_k^j \leq T_i\} \right) p_{\mathbf{G},i}^{j,\mathbf{X}} \right]^{1-\Delta_i} \right\} \tag{5.6}$$

with respect to the β_j, γ_j, and $\Lambda_j\{t_k^j\}$. We now show that such a maximizer exists.

Assume first that $\Lambda_j\{t_k^j\} \leq L < \infty$ for every $k = 1, \ldots, |\mathcal{S}_n^j|$ and $j \in \mathcal{J}$. L_n is a continuous function of the β_j, γ_j, and $\Lambda_j\{t_k^j\}$ on the compact set $\mathcal{B} \times \mathcal{G} \times [0, L]^{s_n}$, where $s_n = \sum_{j \in \mathcal{J}} |\mathcal{S}_n^j|$. Therefore L_n achieves its maximum on this set. To show that a maximum exists on the set $\mathcal{B} \times \mathcal{G} \times [0, \infty)^{s_n}$, we show that there exists a finite L such that for all $(\mathbf{B}^L, \mathbf{G}^L, (\Lambda_j^L\{t_k^j\})_{j,k}) \in (\mathcal{B} \times \mathcal{G} \times [0, \infty)^{s_n}) \backslash (\mathcal{B} \times \mathcal{G} \times [0, L]^{s_n})$, there exists a $(\mathbf{B}, \mathbf{G}, (\Lambda_j\{t_k^j\})_{j,k}) \in \mathcal{B} \times \mathcal{G} \times [0, L]^{s_n}$ which has a larger value of L_n. Consider a proof by contradiction. That is, suppose there does not exist such a L. Then for all $L < \infty$, there exists a $(\mathbf{B}^L, \mathbf{G}^L, (\Lambda_j^L\{t_k^j\})_{j,k}) \in (\mathcal{B} \times \mathcal{G} \times [0, \infty)^{s_n}) \backslash (\mathcal{B} \times \mathcal{G} \times [0, L]^{s_n})$ such that for all $(\mathbf{B}, \mathbf{G}, (\Lambda_j\{t_k^j\})_{j,k}) \in \mathcal{B} \times \mathcal{G} \times [0, L]^{s_n}$, $L_n(\mathbf{B}, \mathbf{G}, (\Lambda_j\{t_k^j\})_{j,k}) \leq L_n(\mathbf{B}^L, \mathbf{G}^L, (\Lambda_j^L\{t_k^j\})_{j,k})$. But we show that $L_n(\mathbf{B}^L, \mathbf{G}^L, (\Lambda_j^L\{t_k^j\})_{j,k})$ can be made arbitrarily small by increasing L, which is a contradiction. To see this, note that (5.6) is bounded from above by

$$J^{n-s_n} \prod_{i=1}^{n} \prod_{j \in \mathcal{J}} \{\Lambda_j\{T_i\} c_3\}^{\Delta_i^j} \exp\left(-c_2 \Delta_i^j \sum_{k=1}^{|\mathcal{S}_n^j|} \Lambda_j\{t_k^j\} 1\{t_k^j \leq T_i\} \right).$$

If $(\mathbf{B}^L, \mathbf{G}^L, (\Lambda_j^L\{t_k^j\})_{j,k}) \in (\mathcal{B} \times \mathcal{G} \times [0, \infty)^{s_n}) \backslash (\mathcal{B} \times \mathcal{G} \times [0, L]^{s_n})$, then there exists at least one $j \in \mathcal{J}$ and one $l \in \{1, \ldots, |\mathcal{S}_n^j|\}$ such that $\Lambda_j^L\{t_l^j\} > L$. Let i^* be the index of the individual such $\Delta_{i^*}^j = 1$ and $T_{i^*} = t_l^j$. Then

$$\left\{\Lambda_j^L\{T_{i^*}\} c_3\right\}^{\Delta_{i^*}^j} \exp\left(-c_2 \Delta_{i^*}^j \sum_{k=1}^{|\mathcal{S}_n^j|} \Lambda_j^L\{t_k^j\} 1\{t_k^j \leq T_{i^*}\} \right)$$

tends to 0 as L tends to $+\infty$. Therefore, the upper bound of $L_n(\mathbf{B}^L, \mathbf{G}^L, (\Lambda_j^L\{t_k^j\})_{j,k})$ can be made as close to 0 as desired by increasing L, which yields a contradiction. Therefore, for any fixed n, the maximum of L_n is obtained in the set $\mathcal{B} \times \mathcal{G} \times [0, L]^{s_n}$, for some $L < \infty$, and on this set, the maximizer $\widehat{\theta}_n$ is achieved.

□

For every n, the problem of maximizing L_n over (5.5) reduces to a finite dimensional one, since the total number s_n of jumps of the N_i $(i = 1, \ldots, n)$ is less than or equal to n.

The expectation-maximization (EM) algorithm (Dempster et al., 1977) can be used to calculate the NPMLEs. Escarela and Bowater (2008) explain and implement to detail the EM algorithm. We briefly explain now,

For $j \in \mathcal{J}$, let $g^j(\mathbf{O};\theta)$ denote the conditional expectation of Γ^j given $\mathbf{O}$ and the parameter value θ. Then $g^j(\mathbf{O};\theta)$ has the form

$$g^j(\mathbf{O};\theta) = \Delta^j + (1-\Delta)w^j(\mathbf{O};\theta),$$

where

$$w^j(\mathbf{O};\theta) = \frac{\exp\left(-\Lambda_j(T)e^{\beta_j'\mathbf{Z}} + \gamma_j'\mathbf{X}\right)}{\sum_{k\in\mathcal{J}}\exp\left(-\Lambda_k(T)e^{\beta_k'\mathbf{Z}} + \gamma_k'\mathbf{X}\right)}.$$

In the M-step of the EM-algorithm, we solve the complete-data score equation conditional on the observed data. In particular, a useful integral equation for $\widehat{\Lambda}_{j,n}$ can be obtained (see Lemma 5.4.1).

Let $\mathbb{P}_n$ denote the empirical probability measure. Then the following holds:

Lemma 5.4.1 *The NPMLE $\widehat{\theta}_n$ satisfies the following equation for every $j \in \mathcal{J}$:*

$$\widehat{\Lambda}_{j,n}(t) = \int_0^t \frac{1}{H_n^j(s;\widehat{\theta}_n)}\, dG_n^j(s), \tag{5.7}$$

where $H_n^j(s;\theta) = \mathbb{P}_n[h^j(s,\mathbf{O};\theta)]$, $h^j(s,\mathbf{O};\theta) = Y(s)e^{\beta_j'\mathbf{Z}}g^j(\mathbf{O};\theta)$, and $G_n^j(s) = \mathbb{P}_n N^j(s)$.

Proof: This result is obtained by following these steps:

1. Taking the derivative with respect to the jump sizes $\Lambda_j\{t_k^j\}$, of the conditional expectation of the complete-data log-likelihood given the observed data and the NPMLE, which is given by

$$l_{\widehat{\theta}_n}(\theta) = \sum_{i=1}^n \sum_{j\in\mathcal{J}} \left\{ \Delta_i^j \sum_{k=1}^{|\mathcal{S}_n^j|} 1\{T_i = t_k^j\}\log\Lambda_j\{t_k^j\} + \Delta_i^j\beta_j'\mathbf{Z}_i \right.$$
$$\left. - g^j(\mathbf{O}_i;\widehat{\theta}_n)e^{\beta_j'\mathbf{Z}_i}\sum_{k=1}^{|\mathcal{S}_n^j|}\Lambda_j\{t_k^j\}1\{t_k^j \le T_i\} + g^j(\mathbf{O}_i;\widehat{\theta}_n)\log p_{\mathbf{G},i}^{j,\mathbf{X}} \right\},$$

2. Setting $(\partial l_{\widehat{\theta}_n}(\theta)/\partial \Lambda_j\{t_k^j\})|_{\theta=\widehat{\theta}_n} = 0$ and solving for $\Lambda_j\{t_k^j\}$.

3. Summing over $\{k \in \{1, \ldots, |\mathcal{S}_n^j|\} : t_k^j \leq t\}$.

□

5.5 Identifiability of the semiparametric mixture model for compering risk.

In this section we study the identifiability property for the semiparametric mixture model for competing risks studied in the previous sections. First, we present the definition of identifiability and of the Kullback-Leibler information. After, we show that the model (5.2)-(5.3) is identifiable.

5.5.1 Definition of identifiability and of the Kullback-Leibler information

Let $\mathcal{P} = \{P_\phi : \phi \in \Phi\}$ be a statistical model where the parameter ϕ can be finite- or infinite-dimensional.

Definition 5.5.1 *A model $\mathcal{P} = \{P_\phi : \phi \in \Phi\}$ is identifiable if the parameterization P_ϕ is one-to-one.*

If the family of probabilities $\mathcal{P}$ can be defined in terms of the family of densities $\mathcal{F}$, i.e $\mathcal{P} = \{P_\phi = (f(y;\phi).\mu) : f \in \mathcal{F}\}$ for a measure μ, then the above identifiability condition is expressed as follows:

$$\forall \phi_1, \phi_2 \in \Phi, f(y;\phi_1) = f(y;\phi_2) \Rightarrow \phi_1 = \phi_2. \tag{5.8}$$

Definition 5.5.2 *Consider two distributions $P_{\phi_1} = f(y,\phi_1).\mu$ and $P_{\phi_2} = f(y,\phi_2).\mu$. The Kullback-Leibler information of P_{ϕ_2} on P_{ϕ_1} is expressed as*

$$K(P_{\phi_1}, P_{\phi_2}) = \int_{\mathcal{Y}} \ln \frac{f(y;\phi_2)}{f(y;\phi_1)} f(y;\phi_2)\mu(dy)$$

The Kullback-Leibler information is not a classical distance because the conditions of symmetry and the triangle inequality is not satisfied. However, this measure translates as the approximation of the probabilities. In Konishi and Kitagawa (2008) and Kullback and Leibler (1951) we can find the following properties of the Kullback-Leibler information,

Proposition 5.5.1 *Let two probabilities P_{ϕ_1} and P_{ϕ_2}. The Kullback-Leibler information has the following properties:*

1. *It is always non-negative: $K(P_{\phi_1}, P_{\phi_2}) \geq 0$.*
2. *$K(P_{\phi_1}, P_{\phi_2}) = 0 \Leftrightarrow P_{\phi_1} = P_{\phi_2}$.*

Proposition 5.5.2 *The parameter ϕ is identifiable if and only if:*

$$\forall \phi_1, \phi_2 \in \Phi, K(P_{\phi_1}, P_{\phi_2}) = 0 \Rightarrow \phi_1 = \phi_2.$$

5.5.2 Identifiability for the semiparametric mixture model for competig risks

In this part, we consider the identifiability of the parameter $\theta = (\mathbf{B}, \mathbf{G}, \Lambda_j; j \in \mathcal{J})$. The result is obtained following the definitions of the Kullback-Leibler information and the identifiability definition.

Proposition 5.5.3 *The model is identifiable that is, if $\theta = (\mathbf{B}, \mathbf{G}, \Lambda_j; j \in \mathcal{J})$ and $\theta^* = (\mathbf{B}^*, \mathbf{G}^*, \Lambda_j^*; j \in \mathcal{J})$ are two elements of Θ, such that $L(\theta_0) = L(\theta^*)$ implies $\theta_0 = \theta^*$.*

Proof: Let be $\theta = (\mathbf{B}, \mathbf{G}, \Lambda_j; j \in \mathcal{J})$ and $\theta^* = (\mathbf{B}^*, \mathbf{G}^*, \Lambda_j^*; j \in \mathcal{J})$ two elements of Θ. If $L(\theta) = L(\theta^*)$ by condition (**C8**) there are a $l \in \mathcal{J}$ such that $\Delta^l = 1$, $y \in [0, \tau]$, $\|\mathbf{z}\| \leq c_1$ and $\|\mathbf{x}\| \leq c_1$, therefore,

$$\lambda_l(y) e^{\beta_l' \mathbf{z}} \exp\left(e^{\beta_l' \mathbf{z}} \Lambda_l(y)\right) p_{\mathbf{G}}^{l,\mathbf{x}} = \lambda_l^*(y) e^{\beta_l^{*'} \mathbf{z}} \exp\left(e^{\beta_l^{*'} \mathbf{z}} \Lambda_l^*(y)\right) p_{\mathbf{G}^*}^{l,\mathbf{x}},$$

which can be rewritten as,

$$p_{\mathbf{G}}^{l,\mathbf{x}} \frac{\partial \exp\left(e^{\beta_l' \mathbf{z}} \Lambda_l(s)\right)}{\partial s} = p_{\mathbf{G}^*}^{l,\mathbf{z}} \frac{\partial \exp\left(e^{\beta_l^{*'} \mathbf{z}} \Lambda_l^*(s)\right)}{\partial s}.$$

Let $t \in [0, \tau]$, integrating both sides of the equation above 0 to t we obtained

$$p_{\mathbf{G}}^{l,\mathbf{x}} \exp\left(e^{\beta_l' \mathbf{z}} \Lambda_l(t))\right) = p_{\mathbf{G}^*}^{l,\mathbf{x}} \exp\left(e^{\beta_l^{*'} \mathbf{z}} \Lambda_l^*(t)\right)$$

equivalently,

$$\frac{\exp\left(e^{\beta_l' \mathbf{z}} \Lambda_l(t))\right)}{\exp\left(e^{\beta_l^{*'} \mathbf{z}} \Lambda_l^*(t)\right)} = \frac{p_{\mathbf{G}^*}^{l,\mathbf{x}}}{p_{\mathbf{G}}^{l,\mathbf{x}}}. \tag{5.9}$$

Note that the right-hand side of the equation (5.9) is independent of t, then

$$\frac{\exp\left(e^{\beta_l' \mathbf{z}} \Lambda_l(t))\right)}{\exp\left(e^{\beta_l^{*'} \mathbf{z}} \Lambda_l^*(t)\right)} = \frac{p_{\mathbf{G}^*}^{l,\mathbf{x}}}{p_{\mathbf{G}}^{l,\mathbf{x}}} = \kappa_1, \tag{5.10}$$

where κ_1 is a positive constant. In particular, taking $\mathbf{x} = 0$ implies that $\frac{p_{\mathbf{G}^*}^{l,\mathbf{x}}}{p_{\mathbf{G}}^{l,\mathbf{x}}} = 1$, i.e., $\kappa_1 = 1$. From this result and the equation (5.10) we obtained

$$\frac{\exp\left(e^{\beta_l' \mathbf{z}} \Lambda_l(t)\right)}{\exp\left(e^{\beta_l^{*'} \mathbf{z}} \Lambda_l^*(t)\right)} = 1,$$

applying the logarithm and rearranging terms, the above equation reduces to

$$\frac{e^{\beta_l' \mathbf{z}}}{e^{\beta_l^{*'} \mathbf{z}}} = \frac{\Lambda_l^*(t)}{\Lambda_l(t)}.$$

Note that the left-hand side of the equation does not depend on t, then

$$\frac{e^{\beta_l' \mathbf{z}}}{e^{\beta_l^{*'} \mathbf{z}}} = \frac{\Lambda_l^*(t)}{\Lambda_l(t)} = \kappa_2 \tag{5.11}$$

with κ_2 is a positive constant. Taking $\mathbf{z} = 0$, then $\kappa_2 = 1$ and therefore

$$\frac{e^{\beta_l' \mathbf{z}}}{e^{\beta_l^{*'} \mathbf{z}}} = 1,$$

which is equivalent to $(\beta_l - \beta_l^*)'\mathbf{z} = 0$ by (**C6**) it follows that $\beta_l - \beta_l^* = 0$. On the other hand, as $\kappa_2 = 1$, by equation (5.11) $\Lambda_l(t) = \Lambda_l^*(t)$. Finally $\gamma = \gamma^*$, this result is shown in Theorem 1 Bettina Grun and Friedrich Leisch in *Identifiability of Finite Mixtures of Multinomial Logit Models with Varying and Fixed Effects* we do not present the proof of this fact in this thesis because the show is long, but the reader can find all the details in this reference.

Therefore, $\theta = \theta^*$, which completes the proof.

□

5.6 Consistency

The purpose of this section is to prove the following result.

Theorem 5.6.1 *Under conditions C1-C9,* $\|\widehat{\mathbf{B}}_n - \mathbf{B}_0\|$, $\|\widehat{\mathbf{G}}_n - \mathbf{G}_0\|$, *and*

$$\sup_{t\in[0,\tau]} |\widehat{\Lambda}_{j,n}(t) - \Lambda_{j,0}(t)|,$$

for every $j \in \mathcal{J}$, *converge to 0 almost surely as* n *tends to infinity.*

The consistency proof is based on techniques developed by Murphy (1994) for the frailty model (see also Chang *et al.* (2005), Kosorok and Song (2007), and Lu (2008) for recent use of these techniques in various other models for right-censored survival data), but the technical details are quite different. Two lemmas are needed before presenting the proof.

Lemma 5.6.1 *For every* $j \in \mathcal{J}$, $\limsup_n \widehat{\Lambda}_{j,n}(\tau) < \infty$ *almost surely.*

Proof: Let $j \in \mathcal{J}$ and $s \in [0,\tau)$. By assumption C6,

$$H_n^j(s;\widehat{\theta}_n) = \mathbb{P}_n[Y(s)e^{\widehat{\beta}_{j,n}'\mathbf{Z}} g^j(\mathbf{O};\widehat{\theta}_n)] \geq c_2 \mathbb{P}_n[Y(s)\Delta^j],$$

thus it follows from the law of large numbers that $H_n^j(s;\widehat{\theta}_n) \geq c_2 P_{\theta_0}[Y(s)\Delta^j] + o(1)$ almost surely. Under the assumptions stated in Section 5.3, $P_{\theta_0}[Y(s)\Delta^j]$ is bounded away from 0. Thus, for every $s \in [0,\tau)$, $H_n^j(s;\widehat{\theta}_n)$ is bounded away from 0 almost surely as n tends to infinity. Moreover, it is easily shown that the jump size of $\widehat{\Lambda}_{j,n}$ at τ is bounded by $1/c_2$. Therefore,

$$0 \leq \widehat{\Lambda}_{j,n}(\tau) \leq O(1)\mathbb{P}_n N^j(\tau-) + \frac{1}{c_2}$$

almost surely as n tends to infinity, which concludes the proof.

□

Lemma 5.6.2 *For every* $j \in \mathcal{J}$ *and* $t \in [0,\tau]$, *define*

$$\widetilde{\Lambda}_{j,n}(t) = \int_0^t \frac{1}{H_n^j(s;\theta_0)}\, dG_n^j(s).$$

Then $\sup_{t\in[0,\tau]} |\widetilde{\Lambda}_{j,n}(t) - \Lambda_{j,0}(t)|$ *converges to 0 almost surely as* n *tends to infinity.*

Proof: We first show that the class of functions $\{h^j(s,\mathbf{O};\theta) : s \in [0,\tau], \theta \in \Theta\}$ is Donsker. Recall that $\Theta = \mathcal{B} \times \mathcal{G} \times \mathcal{L}^{\otimes J}$. In the course of this proof, it will be useful to denote by $\mathcal{B}_j$ and $\mathcal{G}_j$ the parameter space for β_j and γ_j respectively. Consider the class

$$\mathcal{F} = \left\{ w^j(\mathbf{O};\theta) = \frac{\exp\left(-\Lambda_j(T)e^{\beta_j'\mathbf{Z}} + \gamma_j'\mathbf{X}\right)}{\sum_{k\in\mathcal{J}} \exp\left(-\Lambda_k(T)e^{\beta_k'\mathbf{Z}} + \gamma_k'\mathbf{X}\right)} : \theta \in \Theta \right\}. \tag{5.12}$$

Boundedness of $\mathbf{Z}$ and $\mathbf{X}$ and by example 2 from section 4.2 imply that the classes $\{\beta_j'\mathbf{Z} : \beta_j \in \mathcal{B}_j\}$ and $\{\gamma_j'\mathbf{X} : \gamma_j \in \mathcal{G}_j\}$ are Donsker. Differentiability of $e^{\beta_j' Z}$ in Z and boundedness of the derivative imply that $\{e^{\beta_j'\mathbf{Z}} : \beta_j \in \mathcal{B}_j\}$ are Donsker(see example 3 from section 4.2). Moreover, the class of functions mapping T in $\Lambda_j(T)$ indexed by $\Lambda_j \in \mathcal{L}$ is also Donsker (see example 8 from section 4.2). It follows from example 5 from section 4.2 that the class of functions $-\Lambda_j(T)e^{\beta_j'\mathbf{Z}} + \gamma_j' X$ with θ varying over Θ is Donsker, for every $j \in \mathcal{J}$. Then, by example 4 in section 4.2, we conclude that both the numerator and denominator from (5.12) with θ varying over Θ are Donsker classes. Since the denominator is bounded away from 0 imply that $\mathcal{F}$ is Donsker (see example 6 and 7 from section 4.2). By example 6 from section 4.2, $\Delta^j + (1-\Delta)w^j(\mathbf{O};\theta)$ is Donsker as θ ranges over Θ. Finally, $\{Y(s) : s \in [0,\tau]\}$ is Donsker thus, by multiplying Donsker classes, we can conclude that the class $\{h^j(s,\mathbf{O};\theta) : s \in [0,\tau], \theta \in \Theta\}$ is Donsker.

Similar arguments yield that $\{\Delta^j 1\{T \le t\}/P_{\theta_0}\left[h^j(s,\mathbf{O};\theta_0)\right]\mid_{s=T} : t \in [0,\tau]\}$ is also a Donsker class. Next, for every $j \in \mathcal{J}$ and $t \in [0,\tau]$, define

$$\Lambda_j(t;\theta_0) = \int_0^t \frac{1}{P_{\theta_0}[h^j(s,\mathbf{O};\theta_0)]}\, P_{\theta_0} dN^j(s)$$

. Then

$$\begin{aligned}
&\sup_{t\in[0,\tau]} \left|\widetilde{\Lambda}_{j,n}(t) - \Lambda_j(t;\theta_0)\right| \\
&\quad= \sup_{t\in[0,\tau]} \left| \frac{1}{n}\sum_{i=1}^n \frac{\Delta_i^j 1\{T_i \le t\}}{H_n^j(T_i;\theta_0)} - P_{\theta_0}\left[\frac{\Delta^j 1\{T\le t\}}{P_{\theta_0}\left[h^j(s,\mathbf{O};\theta_0)\right]\mid_{s=T}}\right]\right| \\
&\quad\le \sup_{t\in[0,\tau]} \left| \frac{1}{n}\sum_{i=1}^n \Delta_i^j 1\{T_i \le t\} \left\{ \frac{1}{H_n^j(s;\theta_0)} - \frac{1}{P_{\theta_0}\left[h^j(s,\mathbf{O};\theta_0)\right]} \right\}_{\mid_{s=T_i}} \right| \\
&\qquad\qquad + \sup_{t\in[0,\tau]} \left| (\mathbb{P}_n - P_{\theta_0}) \left[\frac{\Delta^j 1\{T\le t\}}{P_{\theta_0}\left[h^j(s,\mathbf{O};\theta_0)\right]\mid_{s=T}}\right]\right| \\
&\quad\le \sup_{s\in[0,\tau]} \left| \frac{1}{H_n^j(s;\theta_0)} - \frac{1}{P_{\theta_0}\left[h^j(s,\mathbf{O};\theta_0)\right]} \right| \\
&\qquad\qquad + \sup_{t\in[0,\tau]} \left| (\mathbb{P}_n - P_{\theta_0}) \left[\frac{\Delta^j 1\{T\le t\}}{P_{\theta_0}\left[h^j(s,\mathbf{O};\theta_0)\right]\mid_{s=T}}\right]\right|.
\end{aligned} \tag{5.13}$$

From the result above, $\{h^j(s,\mathbf{O};\theta_0) : s \in [0,\tau]\}$ is a Donsker and therefore a Glivenko-Cantelli class of functions, and thus $\sup_{s\in[0,\tau]} |H_n^j(s;\theta_0) - P_{\theta_0}[h^j(s,\mathbf{O};\theta_0)]|$ converges to 0 almost surely. Moreover, for every $s \in [0,\tau]$, $P_{\theta_0}[h^j(s,\mathbf{O};\theta_0)] \geq c_2 P_{\theta_0}[Y(s)\Gamma^j]$ (by C6) and thus by assumption C7, $P_{\theta_0}[h^j(s,\mathbf{O};\theta_0)] > 0$ on $[0,\tau]$. Therefore, the first term on the right hand side of (5.13) converges to 0 almost surely. The second term on the right hand side of (5.13) converges almost surely to 0 by the Glivenko-Cantelli property of $\{\Delta^j 1\{T \leq t\}/P_{\theta_0}[h^j(s,\mathbf{O};\theta_0)]\,|_{s=T} : t \in [0,\tau]\}$. Therefore, we conclude that $\widetilde{\Lambda}_{j,n}$ converges uniformly to $\Lambda_j(\cdot;\theta_0)$, almost surely. It is easy to verify that $\Lambda_j(\cdot;\theta_0)$ is equal to $\Lambda_{j,0}$, which concludes the proof.

□

Proof of Theorem 5.6.1: The proof consists of two steps:

(i) To show that every subsequence of n contains a further subsequence where the NPMLE $\widehat{\theta}_n$ converges.

(ii) To show that the set of limits of all convergent subsequences of $\widehat{\theta}_n$ reduces to $\{\theta_0\}$.

Proof of(i). From the compactness of $\mathcal{B}\times\mathcal{G}$, every subsequence of $(\widehat{\mathbf{B}}_n, \widehat{\mathbf{G}}_n)$ has a further subsequence, say $(\widehat{\mathbf{B}}_{\phi(n)}, \widehat{\mathbf{G}}_{\phi(n)})$, which converges to some $(\mathbf{B}^*, \mathbf{G}^*)$ in $\mathcal{B}\times\mathcal{G}$. Let $j \in \mathcal{J}$. By Lemma 5.6.1 and Helly's theorem, we can find with probability 1 a subsequence $\widehat{\Lambda}_{j,\varphi(n)}$ of $\widehat{\Lambda}_{j,\phi(n)}$ and a nondecreasing right-continuous function Λ_j^* such that $\widehat{\Lambda}_{j,\varphi(n)}(t) \to \Lambda_j^*(t)$ for all $t \in [0,\tau]$ where Λ_j^* is continuous; $\widehat{\Lambda}_{j,\varphi(n)}$ is said to converge weakly to Λ_j^*. By extracting successive sub-subsequences, we can find a further subsequence $\xi(n)$ of $\varphi(n)$ in such a way that this weak convergence holds along $\xi(n)$ for every $j \in \mathcal{J}$. We now show that Λ_j^*, for $j \in \mathcal{J}$, is continuous on $[0,\tau]$. Note first that

$$\widehat{\Lambda}_{j,\xi(n)}(t) = \int_0^t \frac{\mathbb{P}_{\xi(n)}[h^j(s,\mathbf{O};\theta_0)]}{\mathbb{P}_{\xi(n)}[h^j(s,\mathbf{O};\widehat{\theta}_{\xi(n)})]}\, d\widetilde{\Lambda}_{j,\xi(n)}(s), \tag{5.14}$$

where $\widetilde{\Lambda}_{j,n}$ is defined in Lemma 5.6.2. It follows from the Glivenko-Cantelli property of $\{h^j(s,\mathbf{O};\theta) : s \in [0,\tau], \theta \in \Theta\}$ that (see the proof of Lemma 5.6.2)

$$\begin{aligned}
&\sup_{s\in[0,\tau]} \left|\mathbb{P}_{\xi(n)}[h^j(s,\mathbf{O};\theta_0)] - P_{\theta_0}[h^j(s,\mathbf{O};\theta_0)]\right| \longrightarrow 0 \quad a.s.,\\
&\sup_{s\in[0,\tau]} \left|\mathbb{P}_{\xi(n)}[h^j(s,\mathbf{O};\widehat{\theta}_{\xi(n)})] - P_{\theta_0}[h^j(s,\mathbf{O};\widehat{\theta}_{\xi(n)})]\right| \longrightarrow 0 \quad a.s..
\end{aligned} \tag{5.15}$$

Additionally, by using the bounded convergence theorem and the facts that $(\widehat{\mathbf{B}}_{\xi(n)}, \widehat{\mathbf{G}}_{\xi(n)})$ converges to $(\mathbf{B}^*, \mathbf{G}^*)$ and $\widehat{\Lambda}_{j,\xi(n)}$ converges weakly to Λ_j^*, we obtain that $P_{\theta_0}[h^j(s,\mathbf{O};\widehat{\theta}_{\xi(n)})]$ converges to $P_{\theta_0}[h^j(s,\mathbf{O};\theta^*)]$ for every $s \in [0,\tau]$, where $\theta^* = (\mathbf{B}^*, \mathbf{G}^*, \Lambda_j^*; j \in \mathcal{J})$. Moreover, under assumption C3, we can show that the derivative of $P_{\theta_0}[h^j(s,\mathbf{O};\widehat{\theta}_{\xi(n)})]$ with

respect to s is uniformly bounded; hence the sequence of functions $P_{\theta_0}[h^j(\cdot, \mathbf{O}; \widehat{\theta}_{\xi(n)})]$ is equicontinuous. By the Arzela-Ascoli theorem, there exists a subsequence of $\xi(n)$, say $\psi(n)$, such that $P_{\theta_0}[h^j(\cdot, \mathbf{O}; \widehat{\theta}_n)]$ converges uniformly to $P_{\theta_0}[h^j(\cdot, \mathbf{O}; \theta^*)]$ in $[0, \tau]$ along this subsequence; we can assume that this subsequence is the same for all $j \in \mathcal{J}$, by the same argument of extraction of subsequences as above. Using the latter result, equation (5.15), and the triangle inequality, we obtain that

$$\frac{d\widehat{\Lambda}_{j,\psi(n)}(t)}{d\widetilde{\Lambda}_{j,\psi(n)}(t)} = \frac{\mathbb{P}_{\psi(n)}[h^j(t, \mathbf{O}; \theta_0)]}{\mathbb{P}_{\psi(n)}[h^j(t, \mathbf{O}; \widehat{\theta}_{\psi(n)})]} \longrightarrow \frac{P_{\theta_0}[h^j(t, \mathbf{O}; \theta_0)]}{P_{\theta_0}[h^j(t, \mathbf{O}; \theta^*)]}$$

uniformly in $t \in [0, \tau]$. By taking the limits on both sides of $\widehat{\Lambda}_{j,\psi(n)}(t)$ in (5.14), we obtain that

$$\Lambda_j^*(t) = \int_0^t \frac{P_{\theta_0}[h^j(s, \mathbf{O}; \theta_0)]}{P_{\theta_0}[h^j(s, \mathbf{O}; \theta^*)]}\, d\Lambda_{j,0}(s).$$

We conclude that Λ_j^* is absolutely continuous with respect to $\Lambda_{j,0}$, so that $\Lambda_j^*(t)$ is differentiable with respect to t, and therefore continuous. A second conclusion, arising from Dini's theorem, is that $\widehat{\Lambda}_{j,\psi(n)}$ converges uniformly to Λ_j^* with probability 1. In addition, $d\widehat{\Lambda}_{j,\psi(n)}(t)/d\widetilde{\Lambda}_{j,\psi(n)}(t)$ converges to $d\Lambda_j^*(t)/d\Lambda_{j,0}(t) := \lambda_j^*(t)/\lambda_{j,0}(t)$ uniformly in t.

To summarize: for any given subsequence of n, we have found a further subsequence $\psi(n)$ and an element $(\mathbf{B}^*, \mathbf{G}^*, \Lambda_j^*; j \in \mathcal{J})$ such that $\|\widehat{\mathbf{B}}_{\psi(n)} - \mathbf{B}^*\|$, $\|\widehat{\mathbf{G}}_{\psi(n)} - \mathbf{G}^*\|$, and $\sup_{t\in[0,\tau]} |\widehat{\Lambda}_{j,\psi(n)}(t) - \Lambda_j^*(t)|$, for every $j \in \mathcal{J}$, converge to 0 almost surely.

Proof of (ii). Consider the difference

$$0 \leq \frac{1}{\psi(n)} \log L_{\psi(n)}(\widehat{\mathbf{B}}_{\psi(n)}, \widehat{\mathbf{G}}_{\psi(n)}, \widehat{\Lambda}_{j,\psi(n)}; j \in \mathcal{J}) - \frac{1}{\psi(n)} \log L_{\psi(n)}(\mathbf{B}_0, \mathbf{G}_0, \widetilde{\Lambda}_{j,\psi(n)}; j \in \mathcal{J}).$$

By letting n tend to infinity, we obtain that

$$0 \leq P_{\theta_0} \left[\sum_{j\in\mathcal{J}} \log \left(\frac{\lambda_j^*(T) \exp\left(\beta_j^{*\prime}\mathbf{Z} - e^{\beta_j^{*\prime}\mathbf{Z}} \Lambda_j^*(T)\right) p_{\mathbf{G}^*}^{j,\mathbf{X}}}{\lambda_{j,0}(T) \exp\left(\beta_{j,0}'\mathbf{Z} - e^{\beta_{j,0}'\mathbf{Z}} \Lambda_{j,0}(T)\right) p_{\mathbf{G}_0}^{j,\mathbf{X}}} \right)^{\Delta^j} \right.$$
$$\left. + (1-\Delta) \log \left(\frac{\sum_{j\in\mathcal{J}} \exp\left(-e^{\beta_j^{*\prime}\mathbf{Z}} \Lambda_j^*(T)\right) p_{\mathbf{G}^*}^{j,\mathbf{X}}}{\sum_{j\in\mathcal{J}} \exp\left(-e^{\beta_{j,0}'\mathbf{Z}} \Lambda_{j,0}(T)\right) p_{\mathbf{G}_0}^{j,\mathbf{X}}} \right) \right].$$

Since the right side of this inequality is the negative Kullback-Leibler information, then, $P_{\theta_0}\left[log\frac{L(\theta^*)}{L(\theta_0)}\right] = 0$ and therefore, it follows from the proposition 5.5.3 that $\theta^* = \theta_0$.

Combining the results from steps (i) and (ii), we conclude that the original sequences $\|\widehat{\mathbf{B}}_n - \mathbf{B}_0\|$, $\|\widehat{\mathbf{G}}_n - \mathbf{G}_0\|$, and, for every $j \in \mathcal{J}$, $\sup_{t\in[0,\tau]} |\widehat{\Lambda}_{j,n}(t) - \Lambda_{j,0}(t)|$ converge to 0 almost surely as n tends to infinity.

□

5.7 Asymptotic normality

5.7.1 Score and information

Once the consistency has been proved, we can establish the asymptotic distribution of the NPMLEs in Escarela and Bowater (2008). To derive the asymptotic normality, we adapt the function analytic approach developed by Murphy (1995) for the frailty model; see also Fang et al. (2005), Kosorok and Song (2007), and Lu (2008), who recently adapted this approach to various other semiparametric regression models for survival data.

To calculate the score equations, we work with one-dimensional submodels $\widehat{\theta}_{n,\epsilon}$ passing through the estimator $\widehat{\theta}_n$, and we differentiate with respect to ϵ. Specifically, consider the submodel

$$\epsilon \longmapsto \widehat{\theta}_{n,\epsilon} = \left(\widehat{\mathbf{B}}_n + \epsilon \mathbf{h}_{\mathbf{B}}, \widehat{\mathbf{G}}_n + \epsilon \mathbf{h}_{\mathbf{G}}, \int_0^{\cdot} \left(1 + \epsilon h_{\Lambda_j}(s)\right)\, d\widehat{\Lambda}_{j,n}(s); j \in \mathcal{J} \right),$$

where $\mathbf{h}_{\mathbf{B}} = (h'_{\beta_1} \dots h'_{\beta_J})'$, $\mathbf{h}_{\mathbf{G}} = (h'_{\gamma_1} \dots h'_{\gamma_{J-1}})'$, h_{β_j} is a p-dimensional vector ($j \in \mathcal{J}$), h_{γ_j} is a q-dimensional vector ($j = 1, \dots, J-1$), and h_{Λ_j} is a non-negative function on $[0, \tau]$ ($j \in \mathcal{J}$). Let $\mathbf{h}$ denote the collected $(\mathbf{h}_{\mathbf{B}}, \mathbf{h}_{\mathbf{G}}, h_{\Lambda_j}; j \in \mathcal{J})$.

To obtain the score equations, we differentiate $l_{\widehat{\theta}_n}(\widehat{\theta}_{n,\epsilon})$ with respect to ϵ and we evaluate at $\epsilon = 0$. $\widehat{\theta}_n$ maximizes $l_{\widehat{\theta}_n}(\theta)$ and therefore satisfies

$$\left. \frac{\partial l_{\widehat{\theta}_n}(\widehat{\theta}_{n,\epsilon})}{\partial \epsilon} \right|_{\epsilon=0} = 0 \tag{5.16}$$

for every $\mathbf{h}$. Define

$$\Psi_{\mathbf{B}}(\theta) = (\Psi_{\beta_1}(\theta)' \dots \Psi_{\beta_J}(\theta)')'$$

and

$$\Psi_{\mathbf{G}}(\theta) = (\Psi_{\gamma_1}(\theta)' \dots \Psi_{\gamma_{J-1}}(\theta)')',$$

where, for every $j \in \mathcal{J}$,

$$\Psi_{\beta_j}(\theta) = \Delta^j \mathbf{Z} - g^j(\mathbf{O}, \theta) \mathbf{Z} e^{\beta_j' \mathbf{Z}} \Lambda_j(T),$$

and for every $j = 1, \dots, J-1$,

$$\Psi_{\gamma_j}(\theta) = \mathbf{X} \left(g^j(\mathbf{O}, \theta) - p_{\mathbf{G}}^{j,\mathbf{X}} \right).$$

For every $j \in \mathcal{J}$, define also

$$\Psi_{\Lambda_j}(\theta)(h_{\Lambda_j}) = \Delta^j h_{\Lambda_j}(T) - g^j(\mathbf{O}, \theta) e^{\beta_j' Z} \int_0^T h_{\Lambda_j}(s)\, d\Lambda_j(s).$$

Then, after some simple algebra, the score equation (5.16) can be re-expressed as $\Psi_n(\widehat{\theta}_n)(\mathbf{h}) = 0$, where $\Psi_n(\widehat{\theta}_n)(\mathbf{h})$ has the form

$$\Psi_n(\widehat{\theta}_n)(\mathbf{h}) = \mathbb{P}_n \left[\mathbf{h}_\mathbf{B}' \Psi_\mathbf{B}(\widehat{\theta}_n) + \mathbf{h}_\mathbf{G}' \Psi_\mathbf{G}(\widehat{\theta}_n) + \sum_{j \in \mathcal{J}} \Psi_{\Lambda_j}(\widehat{\theta}_n)(h_{\Lambda_j}) \right]. \tag{5.17}$$

We take the space of elements $\mathbf{h}$ to be

$$\mathcal{H} = \big\{ \mathbf{h} = (\mathbf{h}_\mathbf{B}, \mathbf{h}_\mathbf{G}, h_{\Lambda_j}; j \in \mathcal{J}) : \mathbf{h}_\mathbf{B} \in \mathbb{R}^P, \|\mathbf{h}_\mathbf{B}\| < \infty; \mathbf{h}_\mathbf{G} \in \mathbb{R}^Q, \|\mathbf{h}_\mathbf{G}\| < \infty; \\ h_{\Lambda_j} : [0, \tau] \to \mathbb{R}, \|h_{\Lambda_j}\|_v < \infty, j \in \mathcal{J} \big\},$$

where $\|h_{\Lambda_j}\|_v$ denotes the total variation of h_{Λ_j} on $[0, \tau]$. Furthermore, we take the functions h_{Λ_j} to be continuous from the right at 0. In addition, we define

$$\theta(\mathbf{h}) = \mathbf{h}_\mathbf{B}' \mathbf{B} + \mathbf{h}_\mathbf{G}' \mathbf{G} + \sum_{j \in \mathcal{J}} \int_0^\tau h_{\Lambda_j}(s)\, d\Lambda_j(s),$$

where $\mathbf{h} \in \mathcal{H}$. From this, we can re-consider the parameter θ as a linear functional on $\mathcal{H}$, and the parameter space Θ as a subset of $l^\infty(\mathcal{H})$, which is the space of all bounded real-valued functions on $\mathcal{H}$ whose representation is here given with the uniform norm. Moreover, the score operator Ψ_n appears to be a random map from Θ to the space $l^\infty(\mathcal{H})$.

Remark. Note that appropriate choices for $\mathbf{h}$ allow to extract all components of the original parameter θ; in the present study, we shall denote by 0_r ($r \geq 2$) the r-dimensional column vector having all its components equal to 0.

For example, let $\mathbf{h}_\mathbf{G} = 0_Q$, $h_{\Lambda_j}(\cdot) = 0$ for every $j \in \mathcal{J}$, and let $\mathbf{h}_\mathbf{B} = (h_{\beta_1}' \dots h_{\beta_J}')'$ be such that $h_{\beta_j} = 0_p$ for every $j \in \mathcal{J}$ except for some $j = l$, with h_{β_l} being the p-dimensional vector with a one at the i-th location and zeros elsewhere. This yields the i-th component of β_l.

As another example, let $\mathbf{h}_\mathbf{B} = 0_P$, $\mathbf{h}_\mathbf{G} = 0_Q$, $h_{\Lambda_j}(\cdot) = 0$ for every $j \in \mathcal{J}$ except $h_{\Lambda_l}(\cdot) = 1\{\cdot \leq t\}$, for some $t \in (0, \tau)$. In this case, $\theta(\mathbf{h})$ reduces to $\Lambda_l(t)$.

We now define an "information" operator $\sigma = (\sigma_{\mathbf{B}}, \sigma_{\mathbf{G}}, \sigma_{\Lambda_j}; j \in \mathcal{J}) : \mathcal{H} \to \mathcal{H}$ by

$$\begin{aligned}
\sigma_{\mathbf{B}}(\mathbf{h}) &= P_{\theta_0}\left[2\Psi_{\mathbf{B}}(\theta_0)\sum_{j\in\mathcal{J}}\Delta^j h_{\Lambda_j}(T)\right] + P_{\theta_0}\left[\Psi_{\mathbf{B}}(\theta_0)^{\otimes 2}\right]\mathbf{h}_{\mathbf{B}} \\
&\quad + P_{\theta_0}\left[\Psi_{\mathbf{B}}(\theta_0)\Psi_{\mathbf{G}}(\theta_0)'\right]\mathbf{h}_{\mathbf{G}} \\
\sigma_{\mathbf{G}}(\mathbf{h}) &= P_{\theta_0}\left[2\Psi_{\mathbf{G}}(\theta_0)\sum_{j\in\mathcal{J}}\Delta^j h_{\Lambda_j}(T)\right] + P_{\theta_0}\left[\Psi_{\mathbf{G}}(\theta_0)^{\otimes 2}\right]\mathbf{h}_{\mathbf{G}} \\
&\quad + P_{\theta_0}\left[\Psi_{\mathbf{G}}(\theta_0)\Psi_{\mathbf{B}}(\theta_0)'\right]\mathbf{h}_{\mathbf{B}}
\end{aligned}$$

$$\begin{aligned}
\sigma_{\Lambda_j}(\mathbf{h})(s) &= h_{\Lambda_j}(s)P_{\theta_0}\left[W^j(s,\mathbf{O},\theta_0)\right] \\
&\quad - P_{\theta_0}\left[2\Delta^j h_{\Lambda_j}(T)W^j(s,\mathbf{O},\theta_0) - \left\{W^j(s,\mathbf{O},\theta_0)\right\}^2\int_0^T h_{\Lambda_j}(u)\,d\Lambda_{j,0}(u)\right] \\
&\quad + P_{\theta_0}\left[2W^j(s,\mathbf{O},\theta_0)\sum_{k>j}\left\{W^k(s,\mathbf{O},\theta_0)\int_0^T h_{\Lambda_k}(u)\,d\Lambda_{k,0}(u)\right.\right. \\
&\qquad\qquad \left.\left. - W^k(s,\mathbf{O},\theta_0)\int_0^s h_{\Lambda_k}(u)\,d\Lambda_{k,0}(u) - \Delta^k h_{\Lambda_k}(T)\right\}\right] \\
&\quad - \mathbf{h}_{\mathbf{B}}' P_{\theta_0}\left[2\Psi_{\mathbf{B}}(\theta_0)g^j(\mathbf{O};\theta_0)e^{\beta_{j,0}'\mathbf{Z}}Y(s)\right] \\
&\quad - \mathbf{h}_{\mathbf{G}}' P_{\theta_0}\left[2\Psi_{\mathbf{G}}(\theta_0)g^j(\mathbf{O};\theta_0)e^{\beta_{j,0}'\mathbf{Z}}Y(s)\right],
\end{aligned}$$

where $W^j(s,\mathbf{O},\theta_0) = Y(s)e^{\beta_{j,0}'\mathbf{Z}}g^j(\mathbf{O},\theta_0)$, $j \in \mathcal{J}$, $s \in [0,\tau]$.

Remark. Some of the terms in σ may be simplified by using the properties of the conditional expectation. For example, $P_{\theta_0}[W^j(s,\mathbf{O},\theta_0)]$ in $\sigma_{\Lambda_j}(\mathbf{h})$ simplifies to $P_{\theta_0}[Y(s)e^{\beta_{j,0}'\mathbf{Z}}\Gamma^j]$. However, for variance estimation purposes, we will construct later an empirical version of σ by replacing θ_0 and P_{θ_0} by $\widehat{\theta}_n$ and $\mathbb{P}_n$ respectively in $\sigma_{\mathbf{B}}$, $\sigma_{\mathbf{G}}$, and σ_{Λ_j}. Therefore, it is irrelevant to simplify, for instance, $P_{\theta_0}[W^j(s,\mathbf{O},\theta_0)]$ to $P_{\theta_0}[Y(s)e^{\beta_{j,0}'\mathbf{Z}}\Gamma^j]$, since, for $i = 1,\ldots,n$, some Γ_i^j are missing and, thus, the empirical version of $P_{\theta_0}[Y(s)e^{\beta_{j,0}'\mathbf{Z}}\Gamma^j]$ cannot be calculated.

The following lemmas state some useful properties of the score and information operators.

Lemma 5.7.1 *Let $\mathbf{h} \in \mathcal{H}$. Then $P_{\theta_0}[\Psi_1(\theta_0)(\mathbf{h})] = 0$, and by setting $\sigma_{\mathbf{B}}$, $\sigma_{\mathbf{G}}$, and σ_{Λ_j} $(j \in \mathcal{J})$ as above,*

$$P_{\theta_0}\left[\Psi_1(\theta_0)(\mathbf{h})^2\right] = \mathbf{h}_{\mathbf{B}}'\sigma_{\mathbf{B}}(\mathbf{h}) + \mathbf{h}_{\mathbf{G}}'\sigma_{\mathbf{G}}(\mathbf{h}) + \sum_{j\in\mathcal{J}}\int_0^\tau \sigma_{\Lambda_j}(\mathbf{h})(s)h_{\Lambda_j}(s)\,d\Lambda_{j,0}(s).$$

Proof: Let $j \in \mathcal{J}$. Then

$$\begin{aligned} P_{\theta_0}\left[\Psi_{\beta_j}(\theta_0)\right] &= P_{\theta_0}\left[\Delta^j \mathbf{Z} - g^j(\mathbf{O},\theta_0)\mathbf{Z}e^{\beta'_{j,0}\mathbf{Z}}\Lambda_{j,0}(T)\right] \\ &= P_{\theta_0}\left[\Delta^j \mathbf{Z} - \Gamma^j \mathbf{Z}e^{\beta'_{j,0}\mathbf{Z}}\Lambda_{j,0}(T)\right] \\ &= P_{\theta_0}\left[\mathbf{Z}\Gamma^j M(\tau)\right], \end{aligned}$$

where the second line comes from the properties of the conditional expectation, and $M(t) = N(t) - \sum_{l\in\mathcal{J}} \int_0^t \Gamma^l e^{\beta'_{l,0}\mathbf{Z}} Y(s)\, d\Lambda_{l,0}(s)$ is the counting process martingale with respect to the filtration $\sigma\{N(s), 1\{T \le s, \Delta = 0\}, \mathbf{Z}, \mathbf{X}, H : 0 \le s \le t\}$. $\mathbf{Z}$ and Γ^j are bounded and measurable with respect to the filtration making M a martingale, implying that $P_{\theta_0}\left[\mathbf{Z}\Gamma^j M(\tau)\right] = 0$. Similar arguments imply that $P_{\theta_0}\left[\Psi_{\gamma_j}(\theta_0)\right] = 0$ and $P_{\theta_0}\left[\Psi_{\Lambda_j}(\theta_0)(h_{\Lambda_j})\right] = 0$ for every j. This concludes the first part of the proof. To prove the second result, we develop $\Psi_1(\theta_0)(\mathbf{h})^2$, we obtain

$$\begin{aligned} \Psi_1(\theta_0)(\mathbf{h})^2 &= \left[\mathbf{h}'_{\mathbf{B}}\Psi_{\mathbf{B}}(\theta_0) + \mathbf{h}'_{\mathbf{G}}\Psi_{\mathbf{G}}(\theta_0) + \sum_{j\in\mathcal{J}} \Psi_{\Lambda_j}(\theta_0)(h_{\Lambda_j})\right]^2 \\ &= \mathbf{h}'_{\mathbf{B}}\Psi^2_{\mathbf{B}}(\theta_0)\mathbf{h}_{\mathbf{B}} + \mathbf{h}'_{\mathbf{G}}\Psi^2_{\mathbf{G}}(\theta_0)\mathbf{h}_{\mathbf{G}} + \left(\sum_{j\in\mathcal{J}} \Psi_{\Lambda_j}(\theta_0)(h_{\Lambda_j})\right)^2 \qquad (5.18) \\ &\quad + \mathbf{h}'_{\mathbf{B}}\Psi_{\mathbf{B}}(\theta_0)\Psi_{\mathbf{G}}(\theta_0)'\mathbf{h}_{\mathbf{G}} + \mathbf{h}'_{\mathbf{G}}\Psi_{\mathbf{G}}(\theta_0)\Psi_{\mathbf{B}}(\theta_0)'\mathbf{h}_{\mathbf{B}} \\ &\quad + 2\mathbf{h}'_{\mathbf{B}}\Psi_{\mathbf{B}}(\theta_0)\sum_{j\in\mathcal{J}} \Psi_{\Lambda_j}(\theta_0)(h_{\Lambda_j}) + 2\mathbf{h}'_{\mathbf{G}}\Psi_{\mathbf{G}}(\theta_0)\sum_{j\in\mathcal{J}} \Psi_{\Lambda_j}(\theta_0)(h_{\Lambda_j}) \\ &= \mathbf{h}'_{\mathbf{B}}\Psi^2_{\mathbf{B}}(\theta_0)h_{\mathbf{B}} + \mathbf{h}'_{\mathbf{G}}\Psi^2_{\mathbf{G}}(\theta_0)\mathbf{h}_{\mathbf{G}} + \left(\sum_{j\in\mathcal{J}} \Psi_{\Lambda_j}(\theta_0)(h_{\Lambda_j})\right)^2 \qquad (5.19) \\ &\quad + \mathbf{h}'_{\mathbf{B}}\Psi_{\mathbf{B}}(\theta_0)\Psi_{\mathbf{G}}(\theta_0)'\mathbf{h}_{\mathbf{G}} + \mathbf{h}'_{\mathbf{G}}\Psi_{\mathbf{G}}(\theta_0)\Psi_{\mathbf{B}}(\theta_0)'\mathbf{h}_{\mathbf{B}} \\ &\quad + 2\mathbf{h}'_{\mathbf{B}}\Psi_{\mathbf{B}}(\theta_0)\sum_{j\in\mathcal{J}} \left[\Delta^j h_{\Lambda_j}(T) - g^j(\mathbf{O},\theta)e^{\beta'_j\mathbf{Z}}\int_0^T h_{\Lambda_j}(s)\, d\Lambda_j(s)\right] \\ &\quad + 2\mathbf{h}'_{\mathbf{G}}\Psi_{\mathbf{G}}(\theta_0)\sum_{j\in\mathcal{J}} \left[\Delta^j h_{\Lambda_j}(T) - g^j(\mathbf{O},\theta)e^{\beta'_j\mathbf{Z}}\int_0^T h_{\Lambda_j}(s)\, d\Lambda_j(s)\right] \end{aligned}$$

Note that the term

$$\left(\sum_{j\in\mathcal{J}} \Psi_{\Lambda_j}(\theta_0)(h_{\Lambda_j})\right)^2 = \left(\sum_{j\in\mathcal{J}} \Delta^j h_{\Lambda_j}(T) - g^j(\mathbf{O},\theta)e^{\beta'_j\mathbf{Z}}\int_0^T h_{\Lambda_j}(s)\, d\Lambda_j(s)\right)^2$$

$$
\begin{aligned}
= \sum_{j\in\mathcal{J}} \Bigg(& \left(\Delta^j h_{\Lambda_j}(T)\right)^2 + \left(g^j(\mathbf{O},\theta) e^{\beta_j'\mathbf{Z}} \int_0^T h_{\Lambda_j}(s)\, d\Lambda_j(s) \right)^2 \\
& -2\Delta^j h_{\Lambda_j}(T) g^j(\mathbf{O},\theta) e^{\beta_j'\mathbf{Z}} \int_0^T h_{\Lambda_j}(s)\, d\Lambda_j(s) \\
& +P_{\theta_0}\Bigg[2W^j(s,\mathbf{O},\theta_0) \sum_{k>j} \Bigg\{ W^k(s,\mathbf{O},\theta_0) \int_0^T h_{\Lambda_k}(u)\, d\Lambda_{k,0}(u) \\
& -W^k(s,\mathbf{O},\theta_0) \int_0^s h_{\Lambda_k}(u)\, d\Lambda_{k,0}(u) - \Delta^k h_{\Lambda_k}(T) \Bigg\} \Bigg] \Bigg)
\end{aligned} \tag{5.20}
$$

In the other hand, as $P_{\theta_0}\left[\Psi_{\Lambda_j}(\theta_0)(h_{\Lambda_j})\right] = 0$, that is

$$
P_{\theta_0}\left[\Delta^j h_{\Lambda_j}(T)\right] = P_{\theta_0}\left[g^j(\mathbf{O},\theta) e^{\beta_j'\mathbf{Z}} \int_0^T h_{\Lambda_j}(s)\, d\Lambda_j(s) \right]
$$

for all $\mathbf{h} \in \mathcal{H}$. In paricular, taking $\mathbf{h}^2$ the equality above is satisfied, i.e.

$$
P_{\theta_0}\left[\Delta^j h_{\Lambda_j}^2(T)\right] = P_{\theta_0}\left[g^j(\mathbf{O},\theta) e^{\beta_j'\mathbf{Z}} \int_0^T h_{\Lambda_j}^2(s)\, d\Lambda_j(s) \right] \tag{5.21}
$$

Finally, substituting the results (5.20) and (5.21) in (5.18) and taking the expectation of the resulting expression, we can get the result.

□

Lemma 5.7.2 *The operator σ is one-to-one.*

Proof: Assume $\sigma(\mathbf{h}) = 0$. By Lemma 5.7.1, $P_{\theta_0}\left[\Psi_1(\theta_0)(\mathbf{h})^2\right] = 0$, and therefore $\Psi_1(\theta_0)(\mathbf{h}) = 0$ almost surely.

Let $j \in \mathcal{J}$. By assumption **C8**, for almost every $t \in [0,\tau]$, $\|\mathbf{z}\| < c_1$, and $\|\mathbf{x}\| < c_1$, there exists a non-negligible subset of Ω (say Ω') such that $T(\omega) = t$, $\mathbf{Z}(\omega) = \mathbf{z}$, $\mathbf{X}(\omega) = \mathbf{x}$, $\Delta(\omega) = 1$, and $H(\omega) = j$, when $\omega \in \Omega'$. If the equality $\Psi_1(\theta_0)(\mathbf{h}) = 0$ holds almost surely, then in particular, it holds for some $\omega \in \Omega'$. For such a ω, $\Psi_1(\theta_0)(\mathbf{h}) = 0$ reduces to

$$
h_{\Lambda_j}(t) + h_{\beta_j}'\mathbf{z} + h_{\gamma_j}'\mathbf{x} - \sum_{l=1}^{J-1} h_{\gamma_l}'\mathbf{x} p_{\mathbf{G}_0}^{l,x} - e^{\beta_{j,0}'z}\left[\int_0^t h_{\Lambda_j}(s)\, d\Lambda_{j,0}(s) + h_{\beta_j}'\mathbf{z}\Lambda_{j,0}(t) \right] = 0, \tag{5.22}
$$

with the convention $h_{\gamma_J} = 0$. By choosing t arbitrarily close to 0, (5.22) reduces to

$$
h_{\Lambda_j}(0) + h_{\beta_j}'\mathbf{z} + h_{\gamma_j}'\mathbf{x} - \sum_{l=1}^{J-1} h_{\gamma_l}'\mathbf{x} p_{\mathbf{G}_0}^{l,\mathbf{x}} = 0, \tag{5.23}
$$

since h_{Λ_j} and $\Lambda_{j,0}$ are continuous from the right at 0 and $\Lambda_{j,0}(0) = 0$ (by **C5**). Taking the difference (5.22)-(5.23) yields the following equation for almost all $t \in [0, \tau]$, $\|\mathbf{z}\| < c_1$, and $\|\mathbf{x}\| < c_1$:

$$h_{\Lambda_j}(t) - h_{\Lambda_j}(0) - e^{\beta'_{j,0}\mathbf{z}} \left[\int_0^t h_{\Lambda_j}(s)\, d\Lambda_{j,0}(s) + h'_{\beta_j}\mathbf{z}\Lambda_{j,0}(t) \right] = 0 \tag{5.24}$$

Let $t > 0$. Then $\Lambda_{j,0}(t) > 0$ (by **C5**) and equation (5.24) can be rewritten as

$$\frac{h_{\Lambda_j}(t) - h_{\Lambda_j}(0)}{\Lambda_{j,0}(t)} = e^{\beta'_{j,0}z} \left[r_j(t) + h'_{\beta_j} z \right], \tag{5.25}$$

where $r_j(t) = \int_0^t h_{\Lambda_j}(s)\, d\Lambda_{j,0}(s)/\Lambda_{j,0}(t)$.

Consider first the case where $\beta_{j,0} = 0$. Since the left-hand side of (5.25) does not depend on $\mathbf{z}$, h_{β_j} must equal 0. Next, consider the case where $\beta_{j,0} \neq 0$. Let $t_1, t_2 > 0$. Then $e^{\beta'_{j,0}\mathbf{z}}[r_j(t_1) - r_j(t_2)]$ should not depend on $\mathbf{z}$. By assumption **C6**, the covariance matrix of Z is positive definite, hence we can find two distinct values $\mathbf{z}_1$ and $\mathbf{z}_2$ of $\mathbf{Z}$ such that

$$e^{\beta'_{j,0}\mathbf{z}_1} \left[r_j(t_1) - r_j(t_2) \right] = e^{\beta'_{j,0}\mathbf{z}_2} \left[r_j(t_1) - r_j(t_2) \right].$$

This implies that $r_j(t_1) = r_j(t_2)$, from which we deduce that $h_{\Lambda_j}(t)$ has to be constant (say, equal to c_6) for almost every $t \in (0, \tau]$. From (5.25), we then deduce that $h_{\Lambda_j}(0) = c_6$, which further implies that $h_{\beta_j} = 0$, $c_6 = 0$, and thus $h_{\Lambda_j}(t) = 0$ for almost every $t \in [0, \tau]$. This, together with (5.23) implies that $h_{\gamma_j} = 0$ for every $j = 1, \ldots, J-1$.

By letting j range over $\mathcal{J}$, we conclude that $\mathbf{h_B} = 0$, $\mathbf{h_G} = 0$, and that for every $j \in \mathcal{J}$, $h_{\Lambda_j}(t) = 0$ for almost every $t \in [0, \tau]$.

Putting this in $\sigma_{\Lambda_j}(\mathbf{h})(s) = 0$, we obtain that $h_{\Lambda_j}(s) P_{\theta_0} \left[W^j(s, \mathbf{O}, \theta_0) \right] = 0$ for every $s \in [0, \tau]$ and $j \in \mathcal{J}$. By assumptions **C2**, **C5**, and **C6**, $P_{\theta_0} \left[W^j(\cdot, \mathbf{O}, \theta_0) \right]$ is uniformly bounded away from 0 on $[0, \tau]$. Therefore, h_{Λ_j} is identically equal to 0 on $[0, \tau]$, for every $j \in \mathcal{J}$. We conclude that σ is one-to-one.

□

Lemma 5.7.3 *The operator σ is continuously invertible.*

Proof: Since $\mathcal{H}$ is a Banach space, to prove that σ is continuously invertible, it is sufficient to prove that σ is one-to-one and that it can be written as the sum $A + (\sigma - A)$

of a bounded linear operator A with bounded inverse and a compact operator $\sigma - A$ (Lemma 25.93 of van der Vaart, 1998).

σ is one-to-one by Lemma 5.7.2. Next, define the linear operator $A : \mathcal{H} \to \mathcal{H}$ by $A(\mathbf{h}) = (\mathbf{h_B}, \mathbf{h_G}, h_{\Lambda_j}(\cdot)P_{\theta_0}\left[W^j(\cdot, \mathbf{O}, \theta_0)\right]; j \in \mathcal{J})$. A is bounded (by **C4** and **C6**). In addition, for every $j \in \mathcal{J}$, $P_{\theta_0}\left[W^j(\cdot, \mathbf{O}, \theta_0)\right]$ is uniformly bounded away from 0 on $[0, \tau]$ (by C2, C5, C6). Thus A is invertible with bounded inverse $A^{-1}(\mathbf{h}) = (\mathbf{h_B}, \mathbf{h_G}, h_{\Lambda_j}(\cdot)P_{\theta_0}\left[W^j(\cdot, \mathbf{O}, \theta_0)\right]^{-1}; j \in \mathcal{J})$.

The operator $\sigma - A$ is compact, by using the same techniques as in Lu (2008) we can get the result. Because a bounded linear operator with finite dimensional range is compact, we only need show that the operator $K_{\Lambda_j} : VB(0, \tau) \to VB(0, \tau)$, with $j \in \mathcal{J}$ given by

$$\begin{aligned} K_{\Lambda_j}(h_{\Lambda_j})(s) &= h_{\Lambda_j}(s)P_{\theta_0}\left[W^j(s, \mathbf{O}, \theta_0)\right] \\ &\quad -P_{\theta_0}\left[2\Delta^j h_{\Lambda_j}(T)W^j(s, \mathbf{O}, \theta_0) - \left\{W^j(s, \mathbf{O}, \theta_0)\right\}^2 \int_0^T h_{\Lambda_j}(u)\, d\Lambda_{j,0}(u)\right] \\ &\quad +P_{\theta_0}\left[2W^j(s, \mathbf{O}, \theta_0) \sum_{k>j} \left\{W^k(s, \mathbf{O}, \theta_0) \int_0^T h_{\Lambda_k}(u)\, d\Lambda_{k,0}(u)\right.\right. \\ &\qquad \left.\left. -W^k(s, \mathbf{O}, \theta_0) \int_0^s h_{\Lambda_k}(u)\, d\Lambda_{k,0}(u) - \Delta^k h_{\Lambda_k}(T)\right\}\right] \end{aligned}$$

is compact.

Thus given a sequence of function $h_{\Lambda_j,n}$ with $\|h_{\Lambda_j,n}\|_v \leq 1$, we must show that there exits a subsequence and an element $g \in VB(0, \tau)$ such that $\|K_{\Lambda_j}h_{\Lambda_j,\eta(n)} - g\|_v \to 0$.

Now, note that K_{Λ_j} is a linear operator then, $\|K_{\Lambda_j}h_{\Lambda_j}\|_v \leq M_7 \int |h_{\Lambda_j}(u)| d\Lambda_{j0}(u)$ for every h_{Λ_j} and a fixed constant M_7. Hence it suffices to show that there exits a subsequence $h_{\Lambda_j,\eta(n)}$ of $h_{\Lambda_j,n}$ that converges. Since h_{Λ_j} is of bounded variation, we can write $h_{\Lambda_j,n}$ as the difference of bounded increasing function $h^{(1)}_{\Lambda_j,n}$ and $h^{(2)}_{\Lambda_j,n}$. From Helly's theorem, there exists a subsequence $h^{(1)}_{\Lambda_j,\eta(n)}$ of $h^{(1)}_{\Lambda_j,n}$ which converges pointwise to some $h^{(1)*}_{\Lambda_j}$. There also exists a subsequence $h^{(2)}_{\Lambda_j,\eta(n)}$ of $h^{(2)}_{\Lambda_j,n}$ which converges pointwise to some $h^{(2)*}_{\Lambda_j}$. Then $h_{\Lambda_j,n}$ converge to the difference of the limits by the dominated convergence theorem. It follows that $\sigma - A$ is a compact operator.

□

In the present study, we shall denote the inverse of σ by $\widetilde{\sigma} = (\widetilde{\sigma}_{\mathbf{B}}, \widetilde{\sigma}_{\mathbf{G}}, \widetilde{\sigma}_{\Lambda_j}; j \in \mathcal{J}) : \mathcal{H} \to \mathcal{H}$.

5.7.2 Asymptotic normality of the NPMLEs

We need some further notations to establish asymptotic normality. Let $\{e_1, \ldots, e_P\}$ be the canonical basis of $\mathbb{R}^P$, where e_m is the P-dimensional column vector with a 1 in the m-th position and zeros elsewhere, for every $m = 1, \ldots, P$. We denote by $(u, 0_Q, 0; j \in \mathcal{J})$ the collected vector $\mathbf{h}$ such that $\mathbf{h_B} = u$, $\mathbf{h_G} = 0_Q$, and h_{Λ_j} is identically equal to 0 for every $j \in \mathcal{J}$. Define the linear operator $\varpi : \mathbb{R}^P \to \mathbb{R}^P$ by $u \mapsto \varpi(u) = \widetilde{\sigma}_{\mathbf{B}}((u, 0_Q, 0; j \in \mathcal{J}))$. Here, ϖ is a version of $\widetilde{\sigma}_{\mathbf{B}}$ restricted to be a function of its first argument only, with the other arguments set equal to 0. Also, define the $(P \times P)$ matrix Σ by

$$\Sigma = (\varpi(e_1) \ldots \varpi(e_P)) .$$

Then the following holds:

Theorem 5.7.1 *Under conditions C1-C9, $\sqrt{n}(\widehat{\mathbf{B}}_n - \mathbf{B}_0)$ converges in distribution to a P-variate normal distribution with mean zero and efficient variance Σ.*

Proof of Theorem 5.7.1: Our proof follows the ideas based around the proof of Theorem 3 of Fang *et al.* (2005), but the technical details are substantially different. As

$$\Psi_1^2(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta) = \sqrt{n}\left(\Psi_n(\theta_0) - \Psi(\theta_0)\right)(g) + o_p(1). \tag{5.26}$$

By lemma 5.7.1

$$\begin{aligned} \Psi_1^2(\theta_0)(\mathbf{h})\sqrt{n}(\hat{\theta}_n - \theta) &= \sqrt{n}(\hat{\mathbf{B}}_n - \mathbf{B}_0)'\sigma_{\mathbf{B}}(g) + \sqrt{n}(\hat{\mathbf{G}}_n - \mathbf{G}_0)'\sigma_{\mathbf{G}}(g) \\ &\quad + \sum_{j \in \mathcal{J}} \int_0^\tau \sigma_{\Lambda_j}(g)(s) h_{\Lambda_j}(s)\sqrt{n}\, d(\hat{\Lambda}_n - \Lambda_{j,0})(s). \end{aligned} \tag{5.27}$$

Combining equations (5.26) and (5.27) and the fact that σ is continuously invertible,then we can write $g = \widetilde{\sigma}(\mathbf{h})$, we get that

$$\sqrt{n}\left(\mathbf{h}'_{\mathbf{B}}\left(\widehat{\mathbf{B}}_n - \mathbf{B}_0\right) + \mathbf{h}'_{\mathbf{G}}\left(\widehat{\mathbf{G}}_n - \mathbf{G}_0\right) + \sum_{j \in \mathcal{J}} \int_0^\tau h_{\Lambda_j}(s)\, d\left(\widehat{\Lambda}_{j,n} - \Lambda_{j,0}\right)(s)\right) =$$
$$\sqrt{n}\left(\Psi_n(\theta_0)(\widetilde{\sigma}(\mathbf{h})) - P_{\theta_0}\left[\Psi_1(\theta_0)(\widetilde{\sigma}(\mathbf{h}))\right]\right) + o_p(1).$$

Let $\mathbf{h_G} = 0_Q$ and h_{Λ_j} be identically equal to 0 for every $j \in \mathcal{J}$. The above equation reduces to

$$\sqrt{n}\mathbf{h}'_{\mathbf{B}}\left(\widehat{\mathbf{B}}_n - \mathbf{B}_0\right) = \sqrt{n}\left(\Psi_n(\theta_0)(\widetilde{\sigma}(\breve{\mathbf{h}})) - P_{\theta_0}\left[\Psi_1(\theta_0)(\widetilde{\sigma}(\breve{\mathbf{h}}))\right]\right) + o_p(1), \tag{5.28}$$

where $\breve{\mathbf{h}} = (\mathbf{h_B}, 0_Q, 0; j \in \mathcal{J})$. By the central limit theorem and Lemma 5.7.1, $\sqrt{n}\mathbf{h}'_{\mathbf{B}}(\widehat{\mathbf{B}}_n - \mathbf{B}_0)$ converges in distribution to a normal law with mean zero and variance $P_{\theta_0}[\Psi_1(\theta_0)(\widetilde{\sigma}(\breve{\mathbf{h}}))^2]$,

for every $\mathbf{h_B} \in \mathbb{R}^P$. Now, noting that $\breve{\mathbf{h}} = \sigma(\widetilde{\sigma}(\breve{\mathbf{h}})) = (\sigma_{\mathbf{B}}(\widetilde{\sigma}(\breve{\mathbf{h}})), \sigma_{\mathbf{G}}(\widetilde{\sigma}(\breve{\mathbf{h}})), \sigma_{\Lambda_j}(\widetilde{\sigma}(\breve{\mathbf{h}})); j \in \mathcal{J})$, it follows by Lemma 5.7.1 that

$$P_{\theta_0}\left[\Psi_1(\theta_0)(\widetilde{\sigma}(\breve{\mathbf{h}}))^2\right] = \mathbf{h}'_{\mathbf{B}}\widetilde{\sigma}_{\mathbf{B}}(\breve{\mathbf{h}}) = \mathbf{h}'_{\mathbf{B}}\varpi(\mathbf{h_B}),$$

and thus $P_{\theta_0}[\Psi_1(\theta_0)(\widetilde{\sigma}(\breve{\mathbf{h}}))^2] = \mathbf{h}'_{\mathbf{B}}\Sigma\mathbf{h_B}$. Thus, by the Cramer-Wold device (van der Vaart, 1998), $\sqrt{n}(\widehat{\mathbf{B}}_n - \mathbf{B}_0)$ converges in distribution to a normal distribution with mean zero and variance-covariance matrix Σ.

Next, let $\breve{\mathbf{h}}$ be equal to $\breve{\mathbf{h}}_m = (e_m, 0_Q, 0; j \in \mathcal{J})$ in (5.28), for each $m = 1, \ldots, P$ in turn. This yields the following system of P equations:

$$\sqrt{n}\left(\widehat{\mathbf{B}}_n - \mathbf{B}_0\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} l\left(\mathbf{O}_i; \theta_0\right) + o_p(1),$$

where

$$l\left(\mathbf{O}; \theta_0\right) = \Sigma\Psi_{\mathbf{B}}(\theta_0) + \Sigma^*\Psi_{\mathbf{G}}(\theta_0) + \sum_{j\in\mathcal{J}} \Psi_{\Lambda_j}(\theta_0)\left(\Sigma^{**}\right),$$

$$\Sigma^* = \begin{pmatrix} \widetilde{\sigma}_{\mathbf{G}}(\breve{\mathbf{h}}_1)' \\ \vdots \\ \widetilde{\sigma}_{\mathbf{G}}(\breve{\mathbf{h}}_P)' \end{pmatrix}, \qquad \Sigma^{**} = \begin{pmatrix} \widetilde{\sigma}_{\Lambda_j}(\breve{\mathbf{h}}_1) \\ \vdots \\ \widetilde{\sigma}_{\Lambda_j}(\breve{\mathbf{h}}_P) \end{pmatrix},$$

and $\Psi_{\Lambda_j}(\theta_0)$ is applied componentwise to Σ^{**}. Thus, $\widehat{\mathbf{B}}_n$ is an asymptotically linear estimator of $\mathbf{B}_0$, and its influence function belongs to the tangent space spanned by the score functions. It follows that $\widehat{\mathbf{B}}_n$ is semiparametrically efficient (Tsiatis, 2006).

□

Remark. In a competing risks analysis, the regression parameter $\mathbf{B}$ is usually the parameter of interest. We can however also state an asymptotic normality result for the NPMLEs of $\mathbf{G}$ and the Λ_j. Let $\varsigma : \mathbb{R}^Q \to \mathbb{R}^Q$ be defined by $\varsigma(u) = \widetilde{\sigma}_{\mathbf{G}}((0_P, u, 0; j \in \mathcal{J}))$, let $\{f_1, \ldots, f_Q\}$ be the canonical basis of $\mathbb{R}^Q$, and let $\Upsilon = (\varsigma(f_1) \ldots \varsigma(f_Q))$ be the $(Q \times Q)$ matrix of ς with respect to this basis. Also, let $\mathbf{h}_{j,t}$ be the collected vector $(\mathbf{h_B}, \mathbf{h_G}, h_{\Lambda_j}; j \in \mathcal{J})$ such that $\mathbf{h_B} = 0_P$, $\mathbf{h_G} = 0_Q$, $h_{\Lambda_j}(\cdot) = 1\{\cdot \leq t\}$ for some $t \in (0, \tau)$ and $j \in \mathcal{J}$, and h_{Λ_l} is identically equal to 0 for every $l \in \mathcal{J}$, $l \neq j$.

Theorem 5.7.2 *Under conditions **C1**-**C9**, $\sqrt{n}(\widehat{\mathbf{G}}_n - \mathbf{G}_0)$ converges in distribution to a Q-variate normal distribution with mean zero and variance matrix Υ. Moreover, for every $t \in (0, \tau)$ and $j \in \mathcal{J}$, $\sqrt{n}(\widehat{\Lambda}_{j,n}(t) - \Lambda_{j,0}(t))$ converges in distribution to a normal distribution with mean zero and variance $\sigma^2_{j,t} = \int_0^t \widetilde{\sigma}_{\Lambda_j}(\mathbf{h}_{j,t})(s)\, d\Lambda_{j,0}(s)$.*

Proof of Theorem 5.7.2: This result can be proved in a similar fashion to the proof of Theorem 5.7.1.

□

5.8 Variance estimation

We now turn to the issue of estimating the asymptotic variance of $\widehat{\mathbf{B}}_n$. Since estimation of the asymptotic variance of $\widehat{\Lambda}_{j,n}(t)$ is useful to obtain confidence intervals for survival probabilities, we also provide estimators for the asymptotic variances of the $\widehat{\Lambda}_{j,n}(t)$ (an estimator for the asymptotic variance of $\widehat{\mathbf{G}}_n$ is also obtained). We need some further notations. Define the $(P \times P)$, $(P \times Q)$, $(Q \times P)$, and $(Q \times Q)$ matrices $\mathbb{A}_n^{\mathbf{B}}$, $\mathbb{A}_n^{\mathbf{G}}$, $\mathbb{B}_n^{\mathbf{B}}$, and $\mathbb{B}_n^{\mathbf{G}}$ by

$$\begin{aligned}
\mathbb{A}_n^{\mathbf{B}} &= \mathbb{P}_n \left[\Psi_{\mathbf{B}}(\widehat{\theta}_n)^{\otimes 2} \right], \\
\mathbb{B}_n^{\mathbf{G}} &= \mathbb{P}_n \left[\Psi_{\mathbf{G}}(\widehat{\theta}_n)^{\otimes 2} \right], \\
\mathbb{A}_n^{\mathbf{G}} &= \mathbb{P}_n \left[\Psi_{\mathbf{B}}(\widehat{\theta}_n) \Psi_{\mathbf{G}}(\widehat{\theta}_n)' \right] = \left(\mathbb{B}_n^{\mathbf{B}} \right)'.
\end{aligned}$$

Define the $(P \times s_n)$ partitioned matrix

$$\mathbb{A}_n^{\Lambda} = (\mathbb{A}_n^{\Lambda_1} \dots \mathbb{A}_n^{\Lambda_J}),$$

where for every $j \in \mathcal{J}$, $\mathbb{A}_n^{\Lambda_j}$ is the $(P \times |\mathcal{S}_n^j|)$ matrix whose P-dimensional l-th column $(l = 1, \dots, |\mathcal{S}_n^j|)$ is given by

$$\frac{2}{n} \Psi_{\mathbf{B},(j,l)}(\widehat{\theta}_n),$$

where $\Psi_{\mathbf{B},(j,l)}(\widehat{\theta}_n)$ denotes the value of $\Psi_{\mathbf{B}}(\widehat{\theta}_n)$, calculated for the subject i such that $\Delta_i = 1$ and $T_i = t_l^j$, for $j \in \mathcal{J}$ and $l = 1, \dots, |\mathcal{S}_n^j|$. Similarly, define the $(Q \times s_n)$ partitioned matrix

$$\mathbb{B}_n^{\Lambda} = (\mathbb{B}_n^{\Lambda_1} \dots \mathbb{B}_n^{\Lambda_J}),$$

where for every $j \in \mathcal{J}$, $\mathbb{B}_n^{\Lambda_j}$ is the $(Q \times |\mathcal{S}_n^j|)$ matrix whose l-th column $(l = 1, \dots, |\mathcal{S}_n^j|)$ is given by $(2/n)\Psi_{\mathbf{G},(j,l)}(\widehat{\theta}_n)$, with $\Psi_{\mathbf{G},(j,l)}(\widehat{\theta}_n)$ defined similarly as $\Psi_{\mathbf{B},(j,l)}(\widehat{\theta}_n)$. Define the $(s_n \times P)$ and $(s_n \times Q)$ partitioned matrices

$$\mathbb{C}_n^{\mathbf{B}} = \begin{pmatrix} \mathbb{C}_{n,1}^{\mathbf{B}} \\ \vdots \\ \mathbb{C}_{n,J}^{\mathbf{B}} \end{pmatrix} \text{ and } \mathbb{C}_n^{\mathbf{G}} = \begin{pmatrix} \mathbb{C}_{n,1}^{\mathbf{G}} \\ \vdots \\ \mathbb{C}_{n,J}^{\mathbf{G}} \end{pmatrix},$$

where for every $j \in \mathcal{J}$, $\mathbb{C}^{\mathbf{B}}_{n,j}$ is a $(|\mathcal{S}^j_n| \times P)$ matrix with P-dimensional l-th raw ($l = 1, \ldots, |\mathcal{S}^j_n|$) given by

$$-\mathbb{P}_n \left[2\Psi_{\mathbf{B}}(\widehat{\theta}_n)' g^j(\mathbf{O}; \widehat{\theta}_n) e^{\widehat{\beta}'_{j,n}\mathbf{Z}} Y(t^j_l) \right],$$

and $\mathbb{C}^{\mathbf{G}}_{n,j}$ is a $(|\mathcal{S}^j_n| \times Q)$ matrix with Q-dimensional l-th raw given by

$$-\mathbb{P}_n \left[2\Psi_{\mathbf{G}}(\widehat{\theta}_n)' g^j(\mathbf{O}; \widehat{\theta}_n) e^{\widehat{\beta}'_{j,n}\mathbf{Z}} Y(t^j_l) \right].$$

Next, let $\mathbb{C}^{\Lambda}_n$ be a $(s_n \times s_n)$ partitioned matrix with (j,k)-th element ($j \in \mathcal{J}, k \in \mathcal{J}$) the $(|\mathcal{S}^j_n| \times |\mathcal{S}^k_n|)$ sub-matrix $\mathbb{C}^{\Lambda_k}_{n,j}$ defined as follows by its (l,m)-th element:

$$\begin{aligned}
\mathbb{C}^{\Lambda_k}_{n,j}(l,m) \;=\; & 1\{j=k\} \left\{ 1\{l=m\} \mathbb{P}_n \left[W^k(t^k_m, \mathbf{O}, \widehat{\theta}_n) \right] - \frac{2}{n} W^k(t^k_l, \mathbf{O}_{(k,m)}, \widehat{\theta}_n) \right. \\
& \left. + \mathbb{P}_n \left[\left\{ W^k(t^k_l, \mathbf{O}, \widehat{\theta}_n) \right\}^2 \widehat{\Delta\Lambda_{k,n}}(t^k_m) 1\{t^k_m \le T\} \right] \right\} \\
& + 1\{j<k\} \left\{ \mathbb{P}_n \left[2W^j(t^j_l, \mathbf{O}, \widehat{\theta}_n) W^k(t^j_l, \mathbf{O}, \widehat{\theta}_n) \widehat{\Delta\Lambda_{k,n}}(t^k_m) \left\{ 1\{t^k_m \le T\} \right. \right. \right. \\
& \left. \left. \left. - 1\{t^k_m \le t^j_l\} \right\} \right] - \frac{2}{n} W^j(t^j_l, \mathbf{O}_{(k,m)}, \widehat{\theta}_n) \right\}
\end{aligned}$$

for $l = 1, \ldots, |\mathcal{S}^j_n|$, and $m = 1, \ldots, |\mathcal{S}^k_n|$. In the formula for $\mathbb{C}^{\Lambda_k}_{n,j}(l,m)$, $\widehat{\Delta\Lambda_{k,n}}(t)$ denotes the jump size of $\widehat{\Lambda}_{k,n}$ at time t; that is, $\widehat{\Delta\Lambda_{k,n}}(t) = \widehat{\Lambda}_{k,n}(t) - \widehat{\Lambda}_{k,n}(t-)$. Moreover, $\mathbf{O}_{(k,m)}$ denotes the value of $\mathbf{O}$ for the subject i such that $\Delta_i = 1$ and $T_i = t^k_m$.

Define the partitioned matrix

$$\mathbb{D}_n = \begin{pmatrix} \mathbb{A}^{\mathbf{B}}_n & \mathbb{A}^{\mathbf{G}}_n & \mathbb{A}^{\Lambda}_n \\ \mathbb{B}^{\mathbf{B}}_n & \mathbb{B}^{\mathbf{G}}_n & \mathbb{B}^{\Lambda}_n \\ \mathbb{C}^{\mathbf{B}}_n & \mathbb{C}^{\mathbf{G}}_n & \mathbb{C}^{\Lambda}_n \end{pmatrix}$$

and the matrices

$$\begin{aligned}
\Sigma_n = & \left\{ \mathbb{A}^{\mathbf{B}}_n - \mathbb{A}^{\mathbf{G}}_n (\mathbb{B}^{\mathbf{G}}_n)^{-1} \mathbb{B}^{\mathbf{B}}_n - (\mathbb{A}^{\Lambda}_n - \mathbb{A}^{\mathbf{G}}_n (\mathbb{B}^{\mathbf{G}}_n)^{-1} \mathbb{B}^{\Lambda}_n) \right. \\
& \left. \times (\mathbb{C}^{\Lambda}_n - \mathbb{C}^{\mathbf{G}}_n (\mathbb{B}^{\mathbf{G}}_n)^{-1} \mathbb{B}^{\Lambda}_n)^{-1} (\mathbb{C}^{\mathbf{B}}_n - \mathbb{C}^{\mathbf{G}}_n (\mathbb{B}^{\mathbf{G}}_n)^{-1} \mathbb{B}^{\mathbf{B}}_n) \right\}^{-1}, \\
\Upsilon_n = & \left\{ \mathbb{B}^{\mathbf{G}}_n - \mathbb{B}^{\mathbf{B}}_n (\mathbb{A}^{\mathbf{B}}_n)^{-1} \mathbb{A}^{\mathbf{G}}_n - (\mathbb{B}^{\Lambda}_n - \mathbb{B}^{\mathbf{B}}_n (\mathbb{A}^{\mathbf{B}}_n)^{-1} \mathbb{A}^{\Lambda}_n) \right. \\
& \left. \times (\mathbb{C}^{\Lambda}_n - \mathbb{C}^{\mathbf{B}}_n (\mathbb{A}^{\mathbf{B}}_n)^{-1} \mathbb{A}^{\Lambda}_n)^{-1} (\mathbb{C}^{\mathbf{G}}_n - \mathbb{C}^{\mathbf{B}}_n (\mathbb{A}^{\mathbf{B}}_n)^{-1} \mathbb{A}^{\mathbf{G}}_n) \right\}^{-1}, \\
\Xi_n = & \left\{ \mathbb{C}^{\Lambda}_n - \mathbb{C}^{\mathbf{B}}_n (\mathbb{A}^{\mathbf{B}}_n)^{-1} \mathbb{A}^{\Lambda}_n - (\mathbb{C}^{\mathbf{G}}_n - \mathbb{C}^{\mathbf{B}}_n (\mathbb{A}^{\mathbf{B}}_n)^{-1} \mathbb{A}^{\mathbf{G}}_n) \right. \\
& \left. \times (\mathbb{B}^{\mathbf{G}}_n - \mathbb{B}^{\mathbf{B}}_n (\mathbb{A}^{\mathbf{B}}_n)^{-1} \mathbb{A}^{\mathbf{G}}_n)^{-1} (\mathbb{B}^{\Lambda}_n - \mathbb{B}^{\mathbf{B}}_n (\mathbb{A}^{\mathbf{B}}_n)^{-1} \mathbb{A}^{\Lambda}_n) \right\}^{-1}.
\end{aligned}$$

Also, for any $t \in (0, \tau)$ and $j \in \mathcal{J}$, define the s_n-dimensional vectors

$$\Phi_{j,t,n} = \left(0'_{l^j_n} \quad \widehat{\Delta\Lambda_{j,n}}(t^j_1) 1\{t^j_1 \le t\} \; \ldots \; \widehat{\Delta\Lambda_{j,n}}(t^j_{|\mathcal{S}^j_n|}) 1\{t^j_{|\mathcal{S}^j_n|} \le t\} \quad 0'_{u^j_n} \right)'$$

and

$$U_{j,t,n} = \left(0'_{l_n^j} \quad 1\{t_1^j \leq t\} \ \ldots \ 1\{t^j_{|\mathcal{S}_n^j|} \leq t\} \quad 0'_{u_n^j}\right)',$$

where $l_n^j = \sum_{k=1}^{j-1} |\mathcal{S}_n^k|$ and $u_n^j = \sum_{k=j+1}^{J} |\mathcal{S}_n^k|$, with $l_n^1 = u_n^J = 0$. Then the following holds:

Theorem 5.8.1 *Under conditions **C1-C9**, the variance estimators Σ_n, Υ_n, and $\sigma^2_{j,t,n} = \Phi'_{j,t,n}\Xi_n U_{j,t,n}$ converge in probability to Σ, Υ, and $\sigma^2_{j,t}$ ($t \in (0,\tau)$, $j \in \mathcal{J}$) respectively.*

Proof of Theorem 5.8.1: The proof of relies on arguments that are now somewhat classical (see for example Parner (1998), Dupuy and Mesbah (2004), Fang *et al.* (2005)).

First, we estimate σ by an empirical version $\sigma_n = (\sigma_{\mathbf{B},n}, \sigma_{\mathbf{G},n}, \sigma_{\Lambda_j,n}; j \in \mathcal{J})$ obtained by replacing θ_0 and P_{θ_0} by $\widehat{\theta}_n$ and $\mathbb{P}_n$ respectively in $\sigma_{\mathbf{B}}$, $\sigma_{\mathbf{G}}$, and σ_{Λ_j}. Following the same arguments of Lemma 5.6.2, the functions $\sigma_{\mathbf{B},n}, \sigma_{\mathbf{G},n}, \sigma_{\Lambda_j,n}$ with $j \in \mathcal{J}$ are the Donsker class, then $\|\sigma_n(\mathbf{h}) - \sigma(\mathbf{h})\|_H \to 0$. Since σ_n is invertible with continuous inverse, then we can express $h = \widetilde{\sigma}_n(g)$. Then

$$\begin{aligned} \|\widetilde{\sigma}_n(g) - \widetilde{\sigma}(g)\|_{\mathcal{H}} &= \|\widetilde{\sigma}(\sigma(\mathbf{h})) - \widetilde{\sigma}_n(\sigma_n(\mathbf{h}))\|_{\mathcal{H}} \\ &\leq \sup_{\mathbf{h}\in\mathcal{H}} \frac{\|\widetilde{\sigma}(\mathbf{h})\|_{\mathcal{H}}}{\|\mathbf{h}\|_{\mathcal{H}}} \|\sigma(\mathbf{h}) - \sigma_n(\mathbf{h})\|_{\mathcal{H}}. \end{aligned}$$

As $\|\sigma_n(\mathbf{h}) - \sigma(\mathbf{h})\|_H \to 0$, then $\widetilde{\sigma}_n = (\widetilde{\sigma}_{\mathbf{B},n}, \widetilde{\sigma}_{\mathbf{G},n}, \widetilde{\sigma}_{\Lambda_j,n}; j \in \mathcal{J})$ converge to $\widetilde{\sigma}(\mathbf{h})$ in probability (see Dupuy y Mesbah, 2004).

For every $\mathbf{h_B}$, the asymptotic variance of $\sqrt{n}\mathbf{h}'_{\mathbf{B}}(\widehat{\mathbf{B}}_n - \mathbf{B}_0)$ is $\mathbf{h}'_{\mathbf{B}}\varpi(\mathbf{h_B})$, which is consistently estimated by $\mathbf{h}'_{\mathbf{B}}\widetilde{\sigma}_{\mathbf{B},n}(\breve{\mathbf{h}})$, where $\breve{\mathbf{h}} = (\mathbf{h_B}, 0_Q, 0; j \in \mathcal{J})$. Let

$$\breve{\mathbf{h}}_n = (\breve{\mathbf{h}}_{\mathbf{B},n}, \breve{\mathbf{h}}_{\mathbf{G},n}, \breve{h}_{\Lambda_j,n}; j \in \mathcal{J}) = \widetilde{\sigma}_n(\breve{\mathbf{h}})$$

. Then $\sigma_n(\breve{\mathbf{h}}_n) = \breve{\mathbf{h}}$, which we can write as

$$\begin{cases} \sigma_{\mathbf{B},n}(\breve{\mathbf{h}}_n) = \mathbf{h_B} \\ \sigma_{\mathbf{G},n}(\breve{\mathbf{h}}_n) = 0_Q \\ \sigma_{\Lambda_1,n}(\breve{\mathbf{h}}_n)(s) = 0, \quad s \in [0,\tau] \\ \vdots \\ \sigma_{\Lambda_J,n}(\breve{\mathbf{h}}_n)(s) = 0, \quad s \in [0,\tau]. \end{cases}$$

In particular, let $s = t_1^j, \ldots, t^j_{|\mathcal{S}_n^j|}$ for every $j \in \mathcal{J}$, in the above system. This yields a system of $(P + Q + s_n)$ equations, which we can write in the following matrix form:

$$\mathbb{D}_n \begin{pmatrix} \breve{\mathbf{h}}_{\mathbf{B},n} \\ \breve{\mathbf{h}}_{\mathbf{G},n} \\ \breve{\mathbf{h}}_{\Lambda,n} \end{pmatrix} = \begin{pmatrix} \mathbf{h_B} \\ 0_Q \\ 0_{s_n} \end{pmatrix} \tag{5.29}$$

where $\breve{\mathbf{h}}_{\Lambda,n} = (\breve{h}_{\Lambda_1,n}(t_1^1)\dots\breve{h}_{\Lambda_1,n}(t_{|\mathcal{S}_n^1|}^1)\dots\breve{h}_{\Lambda_J,n}(t_1^J)\dots\breve{h}_{\Lambda_J,n}(t_{|\mathcal{S}_n^J|}^J))'$. Some algebra on (6.32) shows that $\breve{\mathbf{h}}_{\mathbf{B},n} = \Sigma_n\mathbf{h}_{\mathbf{B}}$, with Σ_n as given above and therefore, $\mathbf{h}_{\mathbf{B}}'\Sigma_n\mathbf{h}_{\mathbf{B}}$ is a consistent estimator of the asymptotic variance of $\sqrt{n}\mathbf{h}_{\mathbf{B}}'(\widehat{\mathbf{B}}_n - \mathbf{B}_0)$ for every $\mathbf{h}_{\mathbf{B}}$. It follows that Σ_n is a consistent estimator of Σ. The consistency of Υ_n proceeds similarly, it is thus omitted.

Let $t \in (0,\tau)$ and $j \in \mathcal{J}$. It follows from the dominated convergence theorem and the consistency of $\widetilde{\sigma}_n$ that $\sigma_{j,t,n}^2 = \int_0^t \widetilde{\sigma}_{\Lambda_j,n}(\mathbf{h}_{j,t})(s)\, d\widehat{\Lambda}_{j,n}(s)$ converges in probability to $\sigma_{j,t}^2$; here, $\mathbf{h}_{j,t}$ is as given in Remark 4. Similarly as above, let $\mathbf{h}_n = (\mathbf{h}_{\mathbf{B},n}, \mathbf{h}_{\mathbf{G},n}, h_{\Lambda_j,n}; j \in \mathcal{J}) = \widetilde{\sigma}_n(\mathbf{h}_{j,t})$. Then $\sigma_n(\mathbf{h}_n) = \mathbf{h}_{j,t}$, which we can write as

$$\begin{cases} \sigma_{\mathbf{B},n}(\mathbf{h}_n) = 0_P \\ \sigma_{\mathbf{G},n}(\mathbf{h}_n) = 0_Q \\ \sigma_{\Lambda_j,n}(\mathbf{h}_n)(s) = 1\{s \le t\}, \quad s \in [0,\tau] \\ \sigma_{\Lambda_l,n}(\mathbf{h}_n)(s) = 0, \quad l \in \mathcal{J}, l \ne j, s \in [0,\tau]. \end{cases} \tag{5.30}$$

In particular, letting $s = t_1^j, \dots, t_{|\mathcal{S}_n^j|}^j$ for every $j \in \mathcal{J}$ in (6.33) yields the system

$$\mathbb{D}_n \begin{pmatrix} \mathbf{h}_{\mathbf{B},n} \\ \mathbf{h}_{\mathbf{G},n} \\ \mathbf{h}_{\Lambda,n} \end{pmatrix} = \begin{pmatrix} 0_P \\ 0_Q \\ U_{j,t,n} \end{pmatrix}$$

where $\mathbf{h}_{\Lambda,n} = (h_{\Lambda_1,n}(t_1^1)\dots h_{\Lambda_1,n}(t_{|\mathcal{S}_n^1|}^1)\dots h_{\Lambda_J,n}(t_1^J)\dots h_{\Lambda_J,n}(t_{|\mathcal{S}_n^J|}^J))'$ and $U_{j,t,n}$ is as defined above. Similar algebra as above shows that $\mathbf{h}_{\Lambda,n} = \Xi_n U_{j,t,n}$. Now, simple calculations show that

$$\begin{aligned} \sigma_{j,t,n}^2 &= \int_0^t \widetilde{\sigma}_{\Lambda_j,n}(\mathbf{h}_{j,t})(s)\, d\widehat{\Lambda}_{j,n}(s) \\ &= \sum_{l=1}^{|\mathcal{S}_n^j|} \widetilde{\sigma}_{\Lambda_j,n}(\mathbf{h}_{j,t})(t_l^j)\widehat{\Delta\Lambda_{j,n}}(t_l^j)1\{t_l^j \le t\} \\ &= \Phi_{j,t,n}'\mathbf{h}_{\Lambda,n} \end{aligned}$$

and therefore, $\Phi_{j,t,n}'\Xi_n U_{j,t,n}$ is a consistent estimator for $\sigma_{j,t}^2$.

□

5.9 Simulation experiments

Simulation studies present an important statistical tool to investigate the performance, properties and adequacy of statistical models in pre-specified situations. This section shows the performance of semiparametric mixture model for competing risk which has been put forward in the previous sections.

Two simulation studies were performed,, in the first study we considered a sample size $n = 30$, while for the second study, the sample size was $n = 200$. In both simulations, we considered two distinct causes of failure (J=2) and a covariate $\mathbf{Z}$ which was generated independently from the $N(0, 1)$ distribution. The proportional hazards mixture model expressed by the equations (5.2) and (5.3) was employed to fit data using two specifications. The first one uses the conditional survival distribution $S_j(t)$ which follows the product limit estimator described in Escarela and Bowater (2008). They specifies the conditional survival distribution for the j-th risk as

$$S_j(t) = \prod_{m:t_{j,(m)} \leq t} \alpha_{j,m} \qquad j \in \mathcal{J}, \quad m = 1, ..., k_j,$$

where the α's are non-negative parameters. By setting $\alpha_{j0} = 1$ and setting

$$\alpha_{jm} = \frac{S_j(t_{j,(m+1)})}{S_j(t_{j,(m)})}$$

the specification for the form of the PLE to be used is completed. Therefore, the estimated conditional baseline hazard function becomes

$$\hat{S}_j(t) = \exp\left\{ - \sum_{m:t_{j,(m)} < t} \frac{d_{jm}}{\sum_{m \in R_{jm}} g_m^j(\theta) \exp(\boldsymbol{\beta}_j \mathbf{Z}_m)} \right\},$$

where, for failure type j, let $t_{j,(1)} < ... < t_{j,(k_j)}$ denote the distinct uncensored failure times, R_{jl} denote the set of subjects known to be at risk just prior to $t_{j,(l)}$ and $d_{jl}$is the tied uncensored failures from cause j at time $t_{j,(l)}$. The second method uses the exponential model where the conditional survival distribution is specified by $\lambda_j(t) = \exp\{\kappa_j\}$, where $-\infty < \kappa_j < \infty$, $j = 1, 2$.

The survival times in each study were obtained through the inverse transformation method, two cases were considered. First, we simulate survival times taken from exponential distributions for the two causes. We assume that both component hazard functions have the exponential distribution:

$$\lambda_j(t|\mathbf{Z}) = h_j \exp(\beta_j' \mathbf{Z}) \qquad j = 1, 2.$$

Given that an entity belongs to the first component, a sample survival time due to cause 1 was generated according to $\lambda_1(t|\mathbf{Z})$ using the method proposed by Bender *et al.* (2005), where

$$T = -\frac{-\log(U)}{h_1 \exp(\beta_1' \mathbf{Z})},$$

U is a variable following a uniform distribution on the interval from 0 to 1. Similarly, for an entity belonging to the second component, a sample survival time due to cause 2 was

generating according to $\lambda_2(t|\mathbf{Z})$ and using the transform method, the survial time is

$$T = -\frac{-\log(U)}{h_2 \exp(\beta_2'\mathbf{Z})},$$

where U is a variable following a uniform distribution on the interval from 0 to 1. The true values considered to generate survival times for this study were

$$(h_1, \beta_1, h_2, \beta_2) = (0.5, -0.5, 1.0, -1.0).$$

For the second simulation study we assume both the component hazard functionf follow the weibull distributions, i.e.

$$\lambda_i(t|\mathbf{Z}) = h_i \exp(\beta_i'\mathbf{Z})\nu_i t^{\nu_i - 1} \qquad j = 1, 2,$$

where $\nu > 0$ and the corresponding survival time is (see Bender *et al.* 2005)

$$T = \left(-\frac{-\log(U)}{h_i \exp(\beta_i'\mathbf{Z})}\right)^{1/\nu_i} \qquad j = 1, 2,$$

where U is a variable following a uniform distribution on the interval from 0 to 1. In this case, the true parameter values that we considered

$$(h_1, \beta_1, \nu_1, h_2, \beta_2, \nu_2) = (0.5, -0.5, 0.5, 1.0, -1.0, 1.5).$$

Note that, with these parameter values, the first hazard function corresponds to a decreasing function, whereas the second hazard function corresponds to an increasing function.

In both simulation studies, the parameter vector in the logistic model given by the equation (5.3) was $\mathbf{X} = (1, \mathbf{Z})$ and the true parameter values were $\gamma = (-1.0, 0.5)$. Also, for each entry the censoring time was generated from a uniform distribution $U(d_1, d_2)$ where d_1 and d_2 are some constants. If the j-th failure time were greater than the j-th censoring time, it was taken to be censored at this censoring time. In this study, we considered three different set of values for d_1 and d_2 so that the comparison under different levels of censoring could be investigated. For each simulation set, we generate 100 independent sample and and fitted the simulated data using the product limit estimator and the exponential model.

In the Tables 5.1 and 5.2 we present the average of estimates of coefficients and their standard errors from two methods (product limit estimator and the exponential model) when the sample size is $n = 30$. The Table **??** shows the results obtained when the survival times follow the exponential model. Note that when the censoring level is low, the estimates of the model semiparametric and the parametric are good and note that if the censoring level increases in this case, 23.3 % and 39.8 % both models have poor

estimators. The Table **??** shows the results obtained when the survival times follow the weibull model. For this case, the estimates for both models and for different levels of censorship were wrong. In the case of the semiparametric approach where the censoring level is high it was impossible to obtain the estimates, the program was left running for several hours (over 8) without any response, the problem is that few data with many censored observations and perhaps the data disribucin also causes more problems.

In the Tables 5.3 and 5.4 we present the average of estimates of coefficients and their standard errors from two methods (product limit estimator and the exponential model) when the sample size is $n = 200$. The Table 5.3 shows the results obtained when the survival times follow the exponential model, it can be seen that the proposed semiparametric method and parametric approach are comparable to each other for mildly and moderately censored samples. For heavily censored samples, the parametric approach provides better estimates of the coefficients in comparison to the semiparametric approach. In the other hand, the Table 5.4 shows the results obtained when the survival times follow the weibull model. From this table, it can be seen that the proposed semiparametric approach provides consistently better estimates of the coefficients respect to the exponential model. As mentioned above, this study takes the distribution of survival times different from an exponential distribution.

In conclusion, with the results of the simulations we can say that when the sample size is 200 the semiparametric approach discussed in this chapter is is theoretically satisfactory but is also computationally intensive. The restricting factor in the use in the EM algorithm is a relatively slow rate of convergence.

Table 5.1: Estimates of the coefficients and their standard errors using product limit estimator and exponential conditional baseline hazard functions with n=30 and the survival times follow the exponential model.

Censoring distribution	**Censored**				**PLE**		**Exponential**	
					Probability Model			
		Parameter	Covariate	True value	Estimate	S.E.	Estimate	S.E
U(2.0,9.0)	8.5%	γ_1	Intercept	-1.0	-1.074	0.181	-1.083	0.206
		γ_1	**X**	0.5	0.539	0.048	0.621	0.074
					Hazard Model			
		β_1	**Z**	-0.5	-0.494	0.055	-0.480	0.320
		β_2	**Z**	-1.0	-0.984	0.105	-0.985	0.054
					Probability Model			
		Parameter	Covariate	True value	Estimate	S.E.	Estimate	S.E.
U(0.5,5.0)	23.3%	γ_1	Intercept	-1	-1.249	0.085	-1.076	0.131
		γ_1	**X**	0.5	0.750	0.180	0.702	0.179
					Hazard Model			
		β_1	**Z**	-0.5	-0.073	0.069	-1.700	0.095
		β_2	**Z**	-1.0	-1.068	0.023	-0.986	0.061
					Probability Model			
		Parameter	Covariate	True value	Estimate	S.E.	Estimate	S.E.
U(0.5,1.8)	39.8%	γ_1	Intercept	-1.0	-2.330	0.423	-2.154	0.130
		γ_1	**X**	0.5	2.718	2.5	3.164	1.957
					Hazard Model			
		β_1	**Z**	-0.5	-5.214	2.15	-7.583	2.338
		β_2	**Z**	-1.0	-0.993	0.204	-0.909	0.056

Table 5.2: Estimates of the coefficients and their standard errors using product limit estimator and exponential conditional baseline hazard functions with n=30 and the survival times follow the weibull model.

Censoring distribution	**Censored**				**PLE**		**Exponential**	
					Probability Model			
		Parameter	Covariate	True value	Estimate	S.E.	Estimate	S.E
U(2.0,19.0)	8.7%	γ_1	Intercept	-1.0	-1.241	0.123	-1.167	0.017
		γ_1	**X**	0.5	0.259	0.124	0.562	0.032
					Hazard Model			
		β_1	**Z**	-0.5	-1.375	0.204	-0.834	0.043
		β_2	**Z**	-1.0	-1.110	0.077	-0.738	0.262
					Probability Model			
		Parameter	Covariate	True value	Estimate	S.E.	Estimate	S.E.
U(0.5,5.7)	23.7%	γ_1	Intercept	-1.0	-6.131	2.519	-1.401	0.052
		γ_1	**X**	0.5	2.613	0.925	0.363	0.325
					Hazard Model			
		β_1	**Z**	-0.5	-5.052	2.103	-3.959	1.612
		β_2	**Z**	-1.0	-1.138	0.402	-0.820	0.136
					Probability Model			
		Parameter	Covariate	True Value	Estimate	S.E.	Estimate	S.E.
U(0.5,2)	39.0%	γ_1	Intercept	-1.0	-	-	-9.411	0.688
		γ_1	**X**	0.5	-	-	3.692	0.608
					Hazard Model			
		β_1	**Z**	-0.5	-	-	-6.842	1.45
		β_2	**Z**	-1.0	-	-	-0.864	0.155

Table 5.3: Estimates of the coefficients and their standard errors using product limit estimator and exponential conditional baseline hazard functions with n=200 and the survival times follow the exponential model.

Censoring distribution	**Censored**				**PLE**		**Exponential**	
					Probability Model			
		Parameter	Covariate	True value	Estimate	S.E.	Estimate	S.E
U(2.0,9.0)	8.9%	γ_1	Intercept	-1.0	-1.051	0.078	-1.040	0.041
		γ_1	**X**	0.5	0.502	0.050	0.508	0.027
					Hazard Model			
		β_1	**Z**	-0.5	-0.518	0.035	-0.492	0.019
		β_2	**Z**	-1.0	-1.004	0.079	-1.00	0.010
					Probability Model			
		Parameter	Covariate	True value	Estimate	S.E.	Estimate	S.E.
U(0.5,5.0)	23.0%	γ_1	Intercept	-1.0	-1.032	0.042	-0.991	0.031
		γ_1	**X**	0.5	0.511	0.015	0.537	0.042
					Hazard Model			
		β_1	**Z**	-0.5	-0.545	0.048	-0.530	0.046
		β_2	**Z**	-1.0	-0.996	0.023	-0.987	0.014
					Probability Model			
		Parameter	Covariate	True value	Estimate	S.E.	Estimate	S.E.
U(0.5,1.8)	40.8%	γ_1	Intercept	-1.0	-1.019	0.022	-1.039	0.045
		γ_1	**X**	0.5	0.611	0.122	0.579	0.011
					Hazard Model			
		β_1	**Z**	-0.5	-0.593	0.098	-0.509	0.030
		β_2	**Z**	-1.0	-1.016	0.034	-0.993	0.007

Table 5.4: Estimates of the coefficients and their standard errors using product limit estimator and exponential conditional baseline hazard functions with n=200 and the survival times follow the weibull model.

Censoring distribution	**Censored**				**PLE**		**Exponential**	
					Probability Model			
		Parameter	Covariate	True value	Estimate	S.E.	Estimate	S.E
U(2.0,19.0)	9%	γ_1	Intercept	-1.0	-1.086	0.095	-1.053	0.063
		γ_1	**X**	0.5	0.415	0.110	0.495	0.018
					Hazard Model			
		β_1	**Z**	-0.5	-0.450	0.051	-0.700	0.036
		β_2	**Z**	-1.0	-1.032	0.037	-0.694	0.033
					Probability Model			
		Parameter	Covariate	True value	Estimate	S.E.	Estimate	S.E.
U(0.5,5.7)	22.3%	γ_1	Intercept	-1.0	-1.125	0.140	-1.354	0.155
		γ_1	**X**	0.5	0.365	0.145	0.248	0.103
					Hazard Model			
		β_1	**Z**	-0.5	-0.428	0.091	-0.376	0.131
		β_2	**Z**	-1.0	-1.042	0.072	-0.834	0.380
					Probability Model			
		Parameter	Covariate	True Value	Estimate	S.E.	Estimate	S.E.
U(0.5,2)	38.7%	γ_1	Intercept	-1.0	-1.155	0.158	-1.844	0.140
		γ_1	**X**	0.5	0.523	0.026	0.248	0.200
					Hazard Model			
		β_1	**Z**	-0.5	-0.560	0.066	-0.414	0.012
		β_2	**Z**	-1.0	-1.039	0.042	-0.900	0.087

Chapter 6

The semiparametric transformation cure models

Introduction

Cure data arise from clinical follow-up studies in which there exists a proportion of subjects in the population who would never experience the event of interest. These subjects are usually referred to as *cured*, while the remaining subjects who are susceptible to the event are referred to as *uncured*.

For instance, in some clinical studies, a substantial proportion of patients who respond favorably to treatment subsequently appear to be free of any signs or symptoms of the disease and may be considered cured, while the remaining patients may eventually relapse. Long-term censored survival usually appear in data. Farewell (1986) and Taylor (1995) provided some typical examples in cancer and radiation research. The goals of such studies include estimation of the cure rate, defined as the proportion of cured subjects in the population, and the failure time distribution of the uncured individuals, adjusting for effects of possible covariates. Applications of cure models can be found in many disciplines, including biomedical sciences, economics, sociology and engineering sciences. Maller and Zhou (1996) provide a list of such applications.

A variety of parametric mixture models have been considered in the literature for survival data with potentially cured patients. A parametric mixture model typically uses a logistic model for the cure rate and a particular parametric distribution is assumed as the failure time distribution of the uncured subjects. The density function $f(t)$ and survival function $S(t)$ for the uncured subjects are derived from this distribution, which may also depend on one or more parameters. Berkson and Gage (1952) used a mixture

exponential distributions and a constant cure fraction to fit survival data from studies of breast cancer and stomach cancer. Farewell (1982, 1986) proposed a Weibull regression for survival and logistic regression for the cure fraction. More discussions of parametric mixture models can be found in Boag (1949), Jones *et al.* (1981), Pack and Morgan (1990), Cantor and Shuster (1992), Maller and Zhou (1992), Sposto *et al.* (1992), Lo *et al.* (1993) and Ghitany *et al.* (1994). More general distributions, such as the extended generalized gamma and the generalized F have been proposed (see Yamaguchi, 1992 and Peng *et al.*, 1998). Parametric methods are parsimonious and easy to interpret. However, they can be sensitive to model misspecification. Furthemore, there is often little physical evidence in a clinical study to suggest and justify a specific parametric model.

The semiparametric logistic mixture model provides a more flexible alternative to parametric methods. Previous work for the semiparametric logistic hazard mixture model has focused on developing point estimation procedures. Kuk and Chen (1992) considered estimation of regression parameters using marginal likelihood method and proposed the so-called proportional hazards cure model in which the proportional hazards regression models (Cox, 1972) is specified in the survival times of susceptible subjects while the logistic regression models is utilized in the cure fraction. However, results from their method depend on a Monte Carlo approximation of the likelihood function involved, which is inconvenient for routine use.

Taylor (1995) employed the Kaplan-Meier survivor estimator to estimate the failure time distribution of uncured patients and the EM algorithm to estimate the coefficients of the logistic model, but this model does not allow for covariates in the failure time distribution of uncured patients. Peng and Dear (2000) and Sy and Taylor (2000) studied a general nonparametric mixture model where the proportional hazards assumption is employed in modeling the effect of covariate on the failure time of patients who are not cured. The EM algorithm, the marginal likelihood approach, and multiple imputations are employed to estimate parameters of interest in the model. Their models extend models and improve estimation methods proposed by other researchers. They extend Cox's proportional hazard regression model by allowing a proportion of event-free patients and investigating covariate effects on that proportion. Also, they establish the theoretical properties of the resulting estimators for the proportional hazards cure model.

Fang *et al.* (2005) consider the inference in a semiparametric logistic/proportional hazard mixture model. They establish existence, consistency and asymptotic normality results for semiparametric maximum likelihood estimator. They also derived consistent variance estimate for both parametric and no parametric components. Lu (2008) proposed nonparametric likelihood approach to estimate the cumulative hazard and the regression parameters and obtained asymptotic properties of resulting estimators.

Recently Lu and Ying (2004) proposed an estimating equations approach for the semiparametric transformation cure models, where the class of linear transformation models

are used for the failure time of susceptibles and the logistic regression is used for the cure fraction. Their approach was motivated by the work of Chen *et al.*(2002) and used the martingale integral representation to construct unbiased estimating equations. The large sample properties of the resulting estimators were also studied. However, the proposed algorithm for solving the equations may not converge and the resulting estimators for the regression parameters are not efficient, even when the model specified is the proportional hazards cure model, i.e. the error term of the linear transformation models follows the extreme value distribution. Semiparametric transformation models have been studied in other complex data, such as clustered failure data (see Zeng *et al.*, 2008), recurrent events (see Zeng and Lin, 2007) and change-point situations (Kosorok and Song, 2007).

The main purpose of this chapter is to generalize the Lu and Ying's model. We propose a general class of semiparametric transformation (see Clayton and Cuzick, 1985; Cuzick, 1988; Bickel *et al.* 1993; Cheng *et al.*, 1995) for the analysis of survival data with long-term survivors. The proposed model has the flexibility to include time-dependent explanatory variables. It combines a logistic regression for the probability of event occurrence with the class of transformation models for the time of occurrence. Included as special cases are the proportional hazards cure model (Farewell 1982; Kuk and Chen, 1992; Sy and Taylor, 2000; Peng and Dear, 2000) and the proportional odds cure model. We establish the asymptotic properties of resulting estimators using thel modern empirical process theory and show that the estimators for the regression parameters are semiparametric efficient. We also derived consistence variance estimators for both the parametric and nonparametric component.

Introduction en Français

Les modèles de survie avec fraction immune permettent l'étude de données où il existe une partie d'individus de l'échantillon qui n'expérimenteront jamais l'évènement d'intérêt. Nous dirons que ces individus sont guéris (ou immunes), tandis que les autres sont susceptibles de connaître l'évènement. Par exemple, dans un essai clinique, une proportion de patients peut répondre favorablement au traitement. Par la suite ces patients ne présentent aucun signe ou symptôme de la maladie, et peuvent être considérés comme guéris, tandis que les patients restants peuvent retomber malades tôt ou tard. Mais pour un individu donné, tant qu'il n'est pas retombé malade, nous ne savons s'il est guéri ou non.

Farewell (1986) et Taylor (1995) fournissent quelques exemples typiques de recherche dans le cancer et les radiations. Les objectifs de telles études incluent l'estimation du taux de guérison, de la distribution du temps de décès des individus susceptibles, en ajustant des effets de covariables. Ce type de problème intervient dans de nombreux domaines: sciences biomédicales, économie, sociologie, sciences de l'ingénierie. Maller et Zhou (1996) fournissent une liste de telles applications.

Une grande variété de modèles paramétriques ont été considérés dans la littérature pour des données de survie avec fraction immune. De tels modèles paramétriques sont décrits par Berkson et Gage (1952), Boag (1949), Jones *et al.* (1981), Pack et Morgan (1990), Cantor et Shuster (1992), Maller et Zhou (1992), Sposto *et al.* (1992), Lo *et al.* (1993) et Ghitany *et al.* (1994). Les méthodes paramétriques sont parcimonieuses et faciles d'interpréter. Cependant, elles peuvent être sensibles aux erreurs de spécification du modèle. De plus, il n'existe souvent dans les données que trop peu d'évidence pour suggérer et justifier un modèle paramétrique. Les modèles de mélanges semi-paramétriques fournissent donc une alternative flexible aux méthodes paramétriques. Kuk et Chen (1992), Taylor (1995), Peng et Dear (2000), Sy et Taylor (2000) ont proposé un modèle semi-paramétrique de durées de vie avec fraction immune, où la distribution du temps d'évènement pour un sujet non guéri est donnée par le modèle de Cox (1972). Ces auteurs ont essentiellement considéré les aspects algorithmiques de l'estimation dans ce type de modèles. Du point de vue théorique cette fois, Fang *et al.* (2005) et Lu (2008) ont considéré l'inférence semi-paramétrique dans un modèle de mélange basé sur la loi logistique et le modèle à risques proportionnels pour les sujets non guéris. Récemment, Lu et Ying (2004) ont proposé des équations d'estimation pour un modèle semi-paramétrique de durées de vie avec fraction immune, lorsque la durée d'évènement est modélisée par un modèle de régression semi-paramétrique de transformation linéaire. Des outils de martingale sont utilisés pour établir les propriétés des estimateurs obtenus. Mais ces estimateurs ne sont pas efficaces.

Le propos principal de ce chapitre est de généraliser le modèle proposé par Lu et Ying (2004) pour l'analyse des données de survie avec fraction immune. Une formulation très

générale du modèle de transformation linéaire est adoptée pour la loi de la durée de vie chez les sujets susceptibles (nous adoptons la formulation proposée dans d'autres contextes par Zeng et Lin (2008, modèle de transformation avec données en clusters), Zeng et Lin (2007, modèle de transformation avec évènements récurrents), Kosorok et Song (2007, modèle de transformation avec rupture)). Ce modèle permet de prendre en compte des covariables dépendant du temps. Il combine une régression logistique pour la probabilité de décè avec la classe de modèles de transformation pour le temps de décè. Il inclut comme cas particuliers les modèles à risques proportionnels avec une fraction immune (Farewell 1982; Kuk and Chen, 1992; Sy and Taylor, 2000; Pe! ng and Dear, 2000) et à odds proportionnels (Bennet, 1983; Murphy *et al.* 1997; Zeng *et al.* 2005 et Martinussen et Scheike 2006). Nous construisons des estimateurs, et établissons leurs propriétés asymptotiques.

6.1 The cure models

In this paper, we consider a semiparametric logistic mixture model, which assumes that the underlying population is a mixture of susceptible and nonsusceptible subjects. Here, we study right-censored survival data with potentially cured patients. All susceptible subjects would eventually experience the event if there were no censoring, while the nonsusceptible ones are immune from the event. Under mixture modeling approach, a decomposition of the event time is given by

$$T = \eta T^* + (1-\eta)\infty$$

where $T^* < \infty$ denotes the failure time of a susceptible subject and η indicates, by the value 1 or 0, whether the sampled subject is susceptible or not. Thus, one can model separately the survival distribution for susceptible individuals and the fraction of susceptible ones.

A model for survival data with cured subjects is specified by the following two components:

$$\lambda(t|\mathbf{Z}) = \lim_{dt\to 0} \frac{\mathbb{P}(t \le T^* < t+dt | T^* \ge t, \mathbf{Z})}{dt} \tag{6.1}$$

$$\pi(\gamma'\mathbf{X}) = \mathbb{P}(\eta = 1|\mathbf{X}, \gamma) = \frac{\exp(\gamma'\mathbf{X})}{1+\exp(\gamma'\mathbf{X})} \tag{6.2}$$

where the first term denotes the hazard function for a susceptible subject and the second term is the cure rate which follows a logistic model. $\mathbf{Z}$ and $\mathbf{X}$ are vectors of covariates in $\mathbb{R}^q$ and $\mathbb{R}^p$ respectively ($\mathbf{Z}$ might depend on time). $\mathbf{Z}$ and $\mathbf{X}$ may share some common time-independent components and $\mathbf{X}$ includes 1 so that γ contains the intercept term. The survival function of T is expressed as,

$$\begin{aligned}
S_T(t|\mathbf{X},\mathbf{Z}) &= \mathbb{P}(T > t|\mathbf{X},\mathbf{Z}) \\
&= \mathbb{P}(T > t, \eta = 1|\mathbf{X},\mathbf{Z}) + \mathbb{P}(T > t, \eta = 0|\mathbf{X},\mathbf{Z}) \\
&= \mathbb{P}(T > t|\eta = 1, \mathbf{X},\mathbf{Z})\mathbb{P}(\eta = 1|\mathbf{X}) + \mathbb{P}(T > t|\eta = 0, \mathbf{X},\mathbf{Z})\mathbb{P}(\eta = 0|\mathbf{X}) \\
&= \mathbb{P}(T^* > t|\mathbf{Z})\pi(\gamma'\mathbf{X}) + \mathbb{P}(\eta = 0|\mathbf{X}) \\
&= \pi(\gamma'\mathbf{X})S_{T^*}(t|\mathbf{Z}) + 1 - \pi(\gamma'\mathbf{X}),
\end{aligned}$$

where $S_{T^*}(t|\mathbf{Z})$ is the survival function of the failure time distribution of uncured patients. The density function

$$\begin{aligned}
f_T(t|\mathbf{X},\mathbf{Z}) &= \frac{d}{dt}\{F_T(t|\mathbf{X},\mathbf{Z})\} \\
&= \frac{d}{dt}\{1 - S_T(t|\mathbf{X},\mathbf{Z})\} \\
&= \frac{d}{dt}\{1 - \pi(\gamma'\mathbf{X})S_{T^*}(t|\mathbf{Z}) - 1 + \pi(\gamma'\mathbf{X})\} \\
&= -\pi(\gamma'\mathbf{X})\frac{d}{dt}S_{T^*}(t|\mathbf{Z}) \\
&= \pi(\gamma'\mathbf{X})f_{T^*}(t|\mathbf{Z}),
\end{aligned}$$

where $f_{T^*}(t|\mathbf{Z})$ is the density function of the failure time distribution of uncured patients.

The density and survival functions of cured patients are set equal to zero and one respectively, for every finite value of t, because cured patients will never experience, for example, a relapse or death due to the disease. Therefore, their failure times can be conveniently defined as infinite.

Suppose that T may be right-censored by a positive random variable C. Let τ denote the total follow-up of the study. Define $Y = \min(T, \min(\tau, C))$ and $\Delta = 1\{T \leq \min(\tau, C)\}$, where $1\{\cdot\}$ denotes the indicator function. Furthermore, we assume that the censoring time C is independent of T and η conditional on $\mathbf{Z}$ and $\mathbf{X}$. The data consist of n independent vectors $(Y_i, \Delta_i, \mathbf{Z}_i, \ \mathbf{X}_i)$ $(i, \ldots, n)$. Based on these data, the usual statistical problem consists in estimating the failure time distribution specified by model (6.1).

Assuming that the marginal distributions of the covariates $\mathbf{Z}$ and $\mathbf{X}$ do not depend on the parameters of the failure time and cure rate distributions, the likelihood function for the mixture cure model, from n iid replicates $(Y_i, \Delta_i, \mathbf{Z}_i, \ \mathbf{X}_i)$, $i, \ldots, n$, is given by

$$\prod_{i=1}^{n} \{f_T(Y_i|\mathbf{X}_i, \mathbf{Z}_i)\}^{\Delta_i} \{S_T(Y_i|\mathbf{X}_i, \mathbf{Z}_i)\}^{(1-\Delta_i)}$$
$$= \prod_{i=1}^{n} \{\pi(\gamma'\mathbf{X}_i) f_{T^*}(Y_i|\mathbf{Z}_i)\}^{\Delta_i} \{1 - \pi(\gamma'\mathbf{X}_i) + \pi(\gamma'\mathbf{X}_i) S_{T^*}(Y_i|\mathbf{X}_i, \mathbf{Z}_i)\}^{(1-\Delta_i)}.$$

The construction of this likelihood under cure model was also derived, for example, by Fang *et al.* (2005) and Lu (2008).

6.2 Transformation models

Transformation models provide a wide class of models for the analysis of censored failure time data. This class includes the proportional hazards model and the proportional odds model as particular cases, as well as many other useful alternatives (see Martinussen and Scheike, 2006). In the past few years, a considerable effort has been made to extend the scope of the transformation model to complex data, such as clustered failure data (see Zeng and Lin, 2008), recurrent events (see Zeng and Lin, 2007) and change-point situations (Kosorok and Song, 2007). Other authors who have studied the transformation models are: Cheng *et al.* (1995), Bagdonavičius and Nikulin (1999), Bagdonavičius and Nikulin (2002), Chen *et al.*, (2002) and Slud and Vonta (2004).

6.2.1 The semiparametric transformation model

Let T^* be some random failure time and $\mathbf{Z}$ be a q-dimensional vector of covariates assumed, first, to be time-independent. The class of linear transformation models relates T^* to $\mathbf{Z}$ via the following equation:

$$h(T^*) = -\beta'\mathbf{Z} + \varepsilon, \tag{6.3}$$

where h is an unknown strictly increasing transformation function, $\beta = (\beta_1, \ldots, \beta_q)'$ is a q-dimensional regression parameter of interest and ε is a random error variable with known distribution function F_ε (ε is assumed independent of $\mathbf{Z}$).

It is convenient to reparametrize the model (6.3) as

$$\Lambda(T^*) = e^{-\beta'\mathbf{Z}} e^{\varepsilon},$$

where $\Lambda(u) = \exp(h(u))$ is a strictly increasing positive unknown function such that $\Lambda(0) = 0$ and $\lim_{u\to\infty} \Lambda(u) = \infty$.

Let $F_{T^*|\mathbf{Z}}$ be the distribution function of T^* given $\mathbf{Z}$. Then

$$\begin{aligned} F_{T^*|\mathbf{Z}}(t) &= \mathbb{P}(T^* \le t|\mathbf{Z}) \\ &= \mathbb{P}(\Lambda(T^*) \le \Lambda(t)|\mathbf{Z}) \\ &= \mathbb{P}(e^{-\beta'\mathbf{Z}} e^{\varepsilon} \le \Lambda(t)|\mathbf{Z}) \\ &= \mathbb{P}(e^{\varepsilon} \le \Lambda(t) e^{\beta'\mathbf{Z}}|\mathbf{Z}) \\ &= F_{e^\varepsilon}(\Lambda(t) e^{\beta'\mathbf{Z}}) \end{aligned}$$

where F_{e^ε} denotes the conditional distribution function of e^ε. Then, the hazard function for T^* given $\mathbf{Z}$ is given by

$$\lambda_{T^*|\mathbf{Z}}(t|\mathbf{Z}) = \frac{dF_{T^*|\mathbf{Z}}(t)/dt}{1 - F_{T^*|\mathbf{Z}}(t)} = \frac{dF_{e^\varepsilon}(\Lambda(t)e^{\beta'\mathbf{Z}})/dt}{1 - F_{e^\varepsilon}(\Lambda(t)e^{\beta'\mathbf{Z}})} = \frac{f_{e^\varepsilon}(\Lambda(t)e^{\beta'\mathbf{Z}})\lambda(t)e^{\beta'\mathbf{Z}}}{1 - F_{e^\varepsilon}(\Lambda(t)e^{\beta'\mathbf{Z}})},$$

where $\lambda(\cdot)$ is the derivative of $\Lambda(\cdot)$ and f_{e^ε} denotes the density function of e^ε. Letting λ_{e^ε} be the hazard function of $\exp(\varepsilon)$, we obtain:

$$\lambda_{T^*|\mathbf{Z}}(t|\mathbf{Z}) = \lambda_{e^\varepsilon}(e^{\beta'\mathbf{Z}}\Lambda(t))e^{\beta'\mathbf{Z}}\lambda(t). \tag{6.4}$$

From (6.4) we deduce the following well-known examples of transformation models:

Example 1. Let ε have the extreme value distribution, that is, $F_\varepsilon(u) = 1 - \exp(-e^u)$. Then $\exp(\varepsilon)$ is distributed as a standard exponential random variable and thus $\lambda_{e^\varepsilon}(u) = 1$, for every $u \ge 0$. It follows that

$$\lambda_{T^*|\mathbf{Z}}(t|\mathbf{Z}) = e^{\beta'\mathbf{Z}}\lambda(t),$$

and (6.4) reduces to the hazard function of a Cox proportional hazards model with baseline hazard rate $\lambda(\cdot)$ and cumulative baseline hazard function $\Lambda(\cdot)$.

Example 2. Let ε have the standard logistic distribution that is $F_\varepsilon(u) = \exp(u)/(1+\exp(u))$. Then $\lambda_{e^\varepsilon}(u) = (1+u)^{-1}$ and (6.4) reduces to

$$\lambda_{T^*|\mathbf{Z}}(t|\mathbf{Z}) = \frac{\lambda(t)}{\Lambda(t) + e^{-\beta'\mathbf{Z}}}.$$

The conditional survival function of T^* given $\mathbf{Z}$ can be expressed as

$$S_{T^*|\mathbf{Z}}(t) = \frac{e^{-\beta'\mathbf{Z}}}{\Lambda(t) + e^{-\beta'\mathbf{Z}}}$$

or equivalently as

$$logit(1 - S_{T^*|\mathbf{Z}}(t)) = \beta'\mathbf{Z} + h(t),$$

which is known as the proportional odds model. Several other examples can be found in Kosorok and Song (2007) and Ma and Kosorok (2005).

Some further generalizations of the transformation model are as follows. Let write F_ε as

$$F_\varepsilon(u) = 1 - G(e^u),$$

where G is a known decreasing function such that $G(0) = 1$ and $G(\infty) = 0$. Then,

$$\begin{aligned}
\mathbb{P}(T^* > t|\mathbf{Z}) &= \mathbb{P}(h(T^*) > h(t)|\mathbf{Z}) = \mathbb{P}(-\beta'\mathbf{Z} + \varepsilon > h(t)|\mathbf{Z}) \\
&= \mathbb{P}(\varepsilon > h(t) + \beta'\mathbf{Z}|\mathbf{Z}) = 1 - F_\varepsilon(h(t) + \beta'\mathbf{Z}) \\
&= G\left(e^{h(t)}e^{\beta'\mathbf{Z}}\right) = G\left(e^{\beta'\mathbf{Z}}\Lambda(t)\right) \\
&= G\left(\int_0^t e^{\beta'\mathbf{Z}} d\Lambda(s)\right).
\end{aligned}$$

This equation can be further extended to accommodate time-dependent covariates in the transformation model. Letting $\tilde{\mathbf{Z}}(t) = \{\mathbf{Z}(s) : 0 \le s \le t\}$ be the history of $\mathbf{Z}(\cdot)$ in the time interval $[0, t]$, several authors consider the transformation model defined by the following conditional survival function

$$\mathbb{P}(T^* > t|\tilde{\mathbf{Z}}(t)) = G\left(\int_0^t e^{\beta'\mathbf{Z}(s)} d\Lambda(s)\right).$$

or equivalently, by the following conditional cumulative hazard function $\Lambda(t|\tilde{\mathbf{Z}}(t))$

$$\Lambda(t|\tilde{\mathbf{Z}}(t)) = -\log G\left(\int_0^t e^{\beta'\mathbf{Z}(s)} d\Lambda(s)\right) \equiv H\left(\int_0^t e^{\beta'\mathbf{Z}(s)} d\Lambda(s)\right). \tag{6.5}$$

Note that H is a known increasing function with $H(0) = 0$ and $H(\infty) = \infty$. In particular, this formulation is adopted by: Kosorok and Song (2007) who study a linear transformation model applied to right-censored survival data with a change point in the regression coefficient based on a covariate threshold; Zeng and Lin (2007) who study the transformation model with random effects for recurrent events; Zeng *et al.* (2008) who study the linear transformation model with random effects for clustered failure times.

Remark. Specifying the function H while leaving the function Λ unspecified in (6.5) is equivalent to specifying the distribution of ε while leaving the function h unspecified in (6.3).

6.3 Notation and model assumptions

We consider a semiparametric linear transformation model with time-dependent covariates and cured fraction, which is specified by the equation (6.2) for the cured fraction and by the class of linear transformation model (6.5) for the failure time of susceptible subjects.

We first state some notations and model assumptions that will be used throughout the chapter. All the random variables are defined on a probability space $(\Omega, \mathcal{C}, \mathbb{P})$.

We assume that the vector of covariates $\mathbf{Z}$ is time-dependent. Let $\tilde{\mathbf{Z}}(t) = \{\mathbf{Z}(s) : 0 \leq s \leq t\}$. With the notations given previously, the data consist of n independent and identically distributed copies $\mathbf{O}_i = (Y_i, \Delta_i, \tilde{\mathbf{Z}}_i(Y_i), \mathbf{X}_i)$, $i = 1, \ldots, n$. Denote by $\theta = (\beta, \gamma, \Lambda)$ the vector of parameters in the model (6.2) - (6.5) and by $\theta_0 = (\beta_0, \gamma_0, \Lambda_0)$ the true value of the parameter.

To establish our results, we need the following regularity assumptions:

(C1) The true values β_0 and γ_0 lie in the interior of known compact sets $\mathcal{B} \subset \mathbb{R}^q$ and $\mathcal{G} \subset \mathbb{R}^p$ respectively.

(C2) The covariate vector $\mathbf{X}$ is bounded (that is, $\|\mathbf{X}\|$ is bounded by a finite constant, where $\|\cdot\|$ denotes the Euclidean norm), with positive definite covariance matrix. $\mathbf{Z}(\cdot)$ is a càglàd process with uniformly bounded variation on $[0, \tau]$. In the sequel, we shall use the following notations: $M_1 = \max_{t\in[0,\tau],\beta\in\mathcal{B}} e^{\beta'\mathbf{Z}(t)}$ and $M_2 = \min_{t\in[0,\tau],\beta\in\mathcal{B}} e^{\beta'\mathbf{Z}(t)}$.

(C3) $\Lambda(\cdot)$ is a strictly increasing positive function on $[0, \tau]$. $\Lambda(\cdot)$ is continuously differentiable.

(C4) $H(\cdot)$ is thrice continuously differentiable on $[0, \infty)$, with $H^{(1)}(u) > 0$ and

$$\sup_{u\geq 0}\{|H^{(k)}(u)|\} < \infty$$

$k = 1, 2, 3$, where $H^{(k)}(\cdot)$ denotes the k-th derivative of $H(\cdot)$.

(C5) Let $\Psi : [0, \infty) \to [0, 1]$ be the function defined by $\Psi(u) = 1 - \exp\{-H(u)\}$. There exists a constant $\rho_0 > 0$ such that

$$\limsup_{x\to\infty}(1+x)^{\rho_0}(1-\Psi(x)) < \infty, \quad \limsup_{x\to\infty}(1+x)^{1+\rho_0}(\Psi^{(1)}(x)) < \infty$$

Under the true value θ_0, the expectation of random variables will noted by P_{θ_0}.

(C6) With probability 1, there exists a positive and finite constant M_3 such that

$$P_{\theta_0}[\Delta | Y, X, \tilde{\mathbf{Z}}(Y)] > M_3,$$

$t \in [0, \tau]$.

(C7) The following identifiability condition holds for every $t \in [0, \tau]$: if there exists a vector $\mu \in \mathbb{R}^q$ and a deterministic function $\alpha_0(t)$ such that $\alpha_0(t) + \mu'\mathbf{Z}(t) = 0$ with probability 1, then $\mu = 0$ and $\alpha_0(t) = 0$.

Under model (6.2) - (6.5) and conditions **C1**-**C7**, the likelihood function for the parameter θ from the observations $\mathbf{O}_i$ $(i, \ldots, n)$ is proportional to

$$\begin{aligned} L_n(\theta) &= \prod_{i=1}^{n} \left\{ \pi(\gamma'\mathbf{X}_i) e^{\beta'\mathbf{Z}_i(Y_i)} \lambda(Y_i) H^{(1)} \left(\int_0^{Y_i} e^{\beta'\mathbf{Z}_i(s)} d\Lambda(s) \right) \right. \\ &\quad \left. \times \exp\left\{ -H\left(\int_0^{Y_i} e^{\beta'\mathbf{Z}_i(s)} d\Lambda(s) \right) \right\} \right\}^{\Delta_i} \\ &\quad \times \left\{ 1 - \pi(\gamma'\mathbf{X}_i) + \pi(\gamma'\mathbf{X}_i) \exp\left\{ -H\left(\int_0^{Y_i} e^{\beta'\mathbf{Z}_i(s)} d\Lambda(s) \right) \right\} \right\}^{(1-\Delta_i)} \end{aligned} \tag{6.6}$$

where for any function $f(\cdot)$, $f^{(1)}(\cdot)$ denotes the derivative of $f(\cdot)$, and $\lambda(\cdot)$ denotes the derivative of $\Lambda(\cdot)$.

6.4 Identifiability

In this part, we consider the identifiability of the parameters.

Proposition 6.4.1 *The model is identifiable that is, $L_1(\theta) = L_1(\theta^*)$ a.s. implies $\theta = \theta^*$.*

Proof: Assume that $L_1(\theta) = L_1(\theta^*)$ a.s. By **C6**, there exists a $\omega \in \Omega$ outside the negligible set where $L_1(\theta)$ might differ from $L_1(\theta^*)$, such that $\Delta(\omega) = 1$. In this case, with probability one,

$$\left\{\pi(\gamma'\mathbf{x})e^{\beta'\mathbf{z}(y)}\lambda(y)\Psi^{(1)}\left(\int_0^y e^{\beta'\mathbf{z}(s)}d\Lambda(s)\right)\right\} = \left\{\pi(\gamma^{*'}\mathbf{x})e^{\beta^{*'}\mathbf{z}(y)}\lambda^*(y)\Psi^{(1)}\left(\int_0^y e^{\beta^{*'}\mathbf{z}(s)}d\Lambda^*(s)\right)\right\}.$$

This equation can also be expressed as:

$$\pi(\gamma'\mathbf{x})\frac{\partial\Psi\left(\int_0^y e^{\beta'\mathbf{z}(s)}d\Lambda(s)\right)}{\partial y} = \pi(\gamma^{*'}\mathbf{x})\frac{\partial\Psi\left(\int_0^y e^{\beta^{*'}\mathbf{z}(s)}d\Lambda^*(s)\right)}{\partial y}.$$

Let $t \in [0, \tau]$. Integrating both sides of this equality from 0 to t yields

$$\pi(\gamma'\mathbf{x})\Psi\left(\int_0^t e^{\beta'\mathbf{z}(s)}d\Lambda(s)\right) = \pi(\gamma^{*'}\mathbf{x})\Psi\left(\int_0^t e^{\beta^{*'}\mathbf{z}(s)}d\Lambda^*(s)\right),$$

and consequently,

$$\frac{\Psi\left(\int_0^t e^{\beta'\mathbf{z}(s)}d\Lambda(s)\right)}{\Psi\left(\int_0^t e^{\beta^{*'}\mathbf{z}(s)}d\Lambda^*(s)\right)} = \frac{\pi(\gamma^{*'}\mathbf{x})}{\pi(\gamma'\mathbf{x})}.$$

Note that the right-hand side of this latter equality is independent of t then,

$$\frac{\Psi\left(\int_0^t e^{\beta'\mathbf{z}(s)}d\Lambda(s)\right)}{\Psi\left(\int_0^t e^{\beta^{*'}\mathbf{z}(s)}d\Lambda^*(s)\right)} = \frac{\pi(\gamma^{*'}\mathbf{x})}{\pi(\gamma'\mathbf{x})} = \kappa,$$

where κ is some positive constant. We need to show that if $\frac{\pi(\gamma^{*'}\mathbf{x})}{\pi(\gamma'\mathbf{x})} = \kappa$ for all $\mathbf{x}$, then $\gamma^* = \gamma$. Indeed, if we take $\mathbf{x} = 0$, then $\kappa = \frac{\pi(0)}{\pi(0)}$, therefore $\kappa = 1$.

Now, we need to show that if $\frac{\pi(\gamma^{*'}\mathbf{x})}{\pi(\gamma'\mathbf{x})} = 1$ for all $\mathbf{x}$, then $\gamma^* = \gamma$. Indeed, $\frac{\pi(\gamma^{*'}\mathbf{x})}{\pi(\gamma'\mathbf{x})} = 1$ implies that

$$\pi(\gamma^{*'}\mathbf{x}) = \pi(\gamma'\mathbf{x}). \tag{6.7}$$

By the definition of $\pi(\cdot)$, the equation (6.7) can be expressed as

$$\mathbb{P}(\eta = 1|\mathbf{x}, \gamma^*) = \mathbb{P}(\eta = 1|\mathbf{x}, \gamma),$$

or equivalently as,

$$1 - \mathbb{P}(\eta = 0|\mathbf{x}, \gamma^*) = 1 - \mathbb{P}(\eta = 0|\mathbf{x}, \gamma),$$

that is,

$$\frac{1}{1 + e^{\gamma^{*'}\mathbf{x}}} = \frac{1}{1 + e^{\gamma'\mathbf{x}}}.$$

By **C2**, it follows that $\gamma = \gamma^*$.

As $\kappa = 1$ then,

$$\frac{\Psi\left(\int_0^t e^{\beta' \mathbf{z}(s)} d\Lambda(s)\right)}{\Psi\left(\int_0^t e^{\beta^{*'} \mathbf{z}(s)} d\Lambda^*(s)\right)} = 1,$$

for all $t \in [0, \tau]$. Thus,

$$\Psi\left(\int_0^t e^{\beta' \mathbf{z}(s)} d\Lambda(s)\right) = \Psi\left(\int_0^t e^{\beta^{*'} \mathbf{z}(s)} d\Lambda^*(s)\right).$$

By definition of $\Psi(\cdot)$,

$$1 - \exp\left\{-H\left(\int_0^t e^{\beta' \mathbf{z}(s)} d\Lambda(s)\right)\right\} = 1 - \exp\left\{-H\left(\int_0^t e^{\beta^{*'} \mathbf{z}(s)} d\Lambda^*(s)\right)\right\},$$

and thus,

$$H\left(\int_0^t e^{\beta' \mathbf{z}(s)} d\Lambda(s)\right) = H\left(\int_0^t e^{\beta^{*'} \mathbf{z}(s)} d\Lambda^*(s)\right).$$

Therefore, by **C4**

$$\int_0^t e^{\beta' \mathbf{z}(s)} d\Lambda(s) = \int_0^t e^{\beta^{*'} \mathbf{z}(s)} d\Lambda^*(s).$$

Taking the Radon-Nikodym derivative of both sides with respect to Λ^* and taking logarithms, we obtain

$$\beta' \mathbf{z}(t) + \log(\lambda(t)) = \beta^{*'} \mathbf{z}(t) + \log(\lambda^*(t)),$$

and finally,

$$(\beta - \beta^*)' \mathbf{z}(t) + \log\left(\frac{\lambda(t)}{\lambda^*(t)}\right) = 0.$$

By condition **C7**, it follows that $\beta = \beta^*$ and $\lambda(t) = \lambda^*(t)$. Therefore the model is identifiable.

□

We now turn to estimation in model (6.2) - (6.5).

6.5 Maximum likelihood estimation

It would seem natural to calculate the maximum likelihood estimator (MLE) of θ_0 by maximizing the likelihood (6.6). However, the maximum of this function is infinity when

the function $\Lambda(\cdot)$ ranges over the class of absolutely continuous functions. In Zeng *et al.* (2008), the authors propose to use the NPML estimation approach, which consists in replacing the original maximization space by an increasing sequence of approximating spaces obtained by letting an estimator of the absolutely continuous $\Lambda(\cdot)$ be an increasing step function on $[0, \tau]$ with jumps at the observed failure times s_i. We assume that the data sample contains k ($k \leq n$) distinct uncensored failure times, which we denote and order as $s_1 < \ldots < s_k$. Then, letting $\Lambda\{t\}$ denote the jump size at t of an increasing step function $\Lambda(\cdot)$, we maximize the function $L_n(\beta, \gamma, \Lambda) =$

$$\prod_{i=1}^{n} \left\{ \pi(\gamma' \mathbf{X}_i) e^{\beta' \mathbf{Z}_i(Y_i)} \Lambda\{Y_i\} H^{(1)} \left(\sum_{j=1}^{k} e^{\beta' \mathbf{Z}_i(s_j)} \Lambda\{s_j\} 1\{s_j \leq Y_i\} \right) \right.$$
$$\times \exp\left\{ -H \left(\sum_{j=1}^{k} e^{\beta' \mathbf{Z}_i(s_j)} \Lambda\{s_j\} 1\{s_j \leq Y_i\} \right) \right\} \Bigg\}^{\Delta_i}$$
$$\times \left\{ 1 - \pi(\gamma' \mathbf{X}_i) + \pi(\gamma' \mathbf{X}_i) \exp\left\{ -H \left(\sum_{j=1}^{k} e^{\beta' \mathbf{Z}_i(s_j)} \Lambda\{s_j\} 1\{s_j \leq Y_i\} \right) \right\} \right\}^{(1-\Delta_i)} \tag{6.8}$$

over the space

$$\Theta_n = \{(\beta, \gamma, \Lambda) : \beta \in \mathcal{B}, \gamma \in \mathcal{G}, \Lambda\{s_j\} \in [0, \infty), j = 1, \ldots, k\}.$$

For any fixed n, we then define the maximum likelihood estimator (MLE) $\hat{\theta}_n$ of θ_0 as the value (if it exists) that maximizes L_n over Θ_n (the maximum likelihood estimator obtained by this procedure is sometimes referred to as a nonparametric MLE (NPMLE) and we shall use this terminology in the sequel).

Proposition 6.5.1 *The maximum likelihood estimator $\hat{\theta}_n = (\hat{\beta}_n, \hat{\gamma}_n, \hat{\Lambda}_n(\cdot))$ exists and is achieved.*

Proof: The proof proceeds by contradiction, and follows an approach which was used, for example, by Fang *et al.* (2005) and Hernández Quintero *et al.* (2009), in various other contexts.

It suffices to show that the function Λ has finite jumps. Assume first that $\Lambda\{s_j\} \leq U < \infty$ for every $j = 1, \ldots, k$. The function L_n is a continuous function of the β, γ, and $\Lambda\{s_j\}$ on the compact set $\mathcal{B} \times \mathcal{G} \times [0, U]^k$. Therefore L_n achieves its maximum on this set.

To show that a maximum exists on the set $\mathcal{B} \times \mathcal{G} \times [0, \infty)^k$, we show that there exists a finite U such that for all $(\beta^U, \gamma^U, \Lambda^U\{s_j\}; j = 1, \ldots, k) \in (\mathcal{B} \times \mathcal{G} \times [0, \infty)^k) \backslash (\mathcal{B} \times \mathcal{G} \times [0, U]^k)$, there exists a $(\beta, \gamma, \Lambda\{s_j\}; j = 1, \ldots, k) \in \mathcal{B} \times \mathcal{G} \times [0, U]^k$ which has a larger value of L_n.

Consider a proof by contradiction. That is, suppose there does not exist such a U. Then for all $U < \infty$, there exists a $(\beta^U, \gamma^U, \Lambda^U\{s_j\}; j = 1, \ldots, k) \in (\mathcal{B} \times \mathcal{G} \times [0, \infty)^k) \backslash (\mathcal{B} \times \mathcal{G} \times [0, U]^k)$ such that for all $(\beta, \gamma, \Lambda\{s_j\}; j = 1, \ldots, k) \in \mathcal{B} \times \mathcal{G} \times [0, U]^k$, $L_n(\beta, \gamma, \Lambda\{s_j\}; j = 1, \ldots, k) \leq L_n(\beta^U, \gamma^U, \Lambda^U\{s_j\}; j = 1, \ldots, k)$.

But we show that $L_n(\beta^U, \gamma^U, \Lambda^U\{s_j\}; j = 1, \ldots, k)$ can be made arbitrarily small by increasing U, which is a contradiction. To see this, note that (6.8) is bounded from above by

$$\prod_{i=1}^{n} \left\{ M_1 \Lambda\{Y_i\} H^{(1)} \left(\sum_{j=1}^{k} e^{\beta' \mathbf{Z}_i(s_j)} \Lambda\{s_j\} 1\{s_j \leq Y_i\} \right) \right.$$
$$\left. \times \exp\left\{ -H \left(\sum_{j=1}^{k} M_2 \Lambda\{s_j\} 1\{s_j \leq Y_i\} \right) \right\} \right\}^{\Delta_i}$$

If $(\beta^U, \gamma^U, \Lambda^U\{s_j\}; j = 1, \ldots, k) \in (\mathcal{B} \times \mathcal{G} \times [0, \infty)^k) \backslash (\mathcal{B} \times \mathcal{G} \times [0, U]^k)$, there exists at least one $l \in \{1, \ldots, k\}$ such that $\Lambda^U\{s_l\} > U$. There also exists one $i^* \in \{1, \ldots, n\}$ such that $\Delta_{i^*} = 1$ and $Y_{i^*} = s_l$. For this individual,

$$\Lambda^U\{Y_{i^*}\} H^{(1)} \left(\sum_{j=1}^{k} e^{\beta' \mathbf{Z}_i(s_j)} \Lambda^U\{s_j\} 1\{s_j \leq Y_{i^*}\} \right)$$
$$\times \exp\left\{ -H \left(\sum_{j=1}^{k} M_2 \Lambda^U\{s_j\} 1\{s_j \leq Y_{i^*}\} \right) \right\}$$

tends to 0 as U (and therefore $\Lambda^U\{Y_{i^*}\}$) tends to $+\infty$, by **C5**. Thus the upper bound of $L_n(\beta^U, \gamma^U, \Lambda^U\{s_j\}; j = 1, \ldots, k)$ can be made as close to 0 as desired by increasing U, which yields a contradiction. It follows that for any fixed n, the maximum of L_n is obtained in the set $\mathcal{B} \times \mathcal{G} \times [0, U]^k$, for some $U < \infty$, and on this set, the maximizer $\widehat{\theta}_n$ is achieved.

□

Let $\mathbb{P}_n$ denote the empirical distribution of the data, and recall that P_{θ_0} denotes the expectation with respect to the true underlying distribution.

Lemma 6.5.1 *The NPMLE $\widehat{\theta}_n$ satisfies the following equation, for every $t \in [0, \tau]$*

$$\hat{\Lambda}_n(t) = \int_0^t \frac{dG_n(u)}{W_n(u; \hat{\theta}_n)} \tag{6.9}$$

where $(1/W_n)(u; \hat{\theta}_n)$ and $G_n(u)$ are non-decreasing functions in u, defined by

$$W_n(u; \theta) = \mathbb{P}_n[w(u, \mathbf{O}; \theta)] \qquad \textit{and} \qquad G_n(u) = \mathbb{P}_n\left[\Delta 1\{Y \leq u\}\right]$$

respectively, where $w(u, \mathbf{O}; \theta) = e^{\beta' \mathbf{Z}(u)} 1\{u \leq Y\} \phi(\mathbf{O}; \theta)$ *and*

$$\begin{aligned} \phi(\mathbf{O}; \theta) &= \frac{(1-\Delta)\pi(\gamma' \boldsymbol{X})\Psi^{(1)}\left(\int_0^Y e^{\beta' \mathbf{Z}(u)} d\Lambda(u)\right)}{1 - \pi(\gamma' \boldsymbol{X}) + \pi(\gamma' \boldsymbol{X})\left(1 - \Psi\left(\int_0^Y e^{\beta' \mathbf{Z}(u)} d\Lambda(u)\right)\right)} \\ &\quad - \frac{\Delta \Psi^{(2)}\left(\int_0^Y e^{\beta' \mathbf{Z}(u)} d\Lambda(u)\right)}{\Psi^{(1)}\left(\int_0^Y e^{\beta' \mathbf{Z}(u)} d\Lambda(u)\right)}. \end{aligned} \tag{6.10}$$

Proof: The proof consists of two steps:

1. Taking the derivative, with respect to the jump sizes $\Lambda\{s_j\}$ $(j = 1, \ldots, k)$, of the log-likelihood (log of equation (6.8))

$$\begin{aligned} l_n(\theta) = \sum_{i=1}^{n} \Delta_i \ln \pi(\gamma' \mathbf{X}_i) + \Delta_i \beta' \mathbf{Z}_i(Y_i) + \Delta_i \sum_{j=1}^{k} 1\{Y_i = s_j\} \ln(\Lambda\{s_j\}) \\ + \Delta_i \ln\left(\Psi^{(1)}\left(\sum_{j=1}^{k} e^{\beta' \mathbf{Z}_i(s_j)} \Lambda\{s_j\} 1\{s_j \leq Y_i\}\right)\right) \\ + (1 - \Delta_i) \ln\left(1 - \pi(\gamma' \mathbf{X}_i) + \pi(\gamma' \mathbf{X}_i)\left(1 - \Psi\left(\sum_{j=1}^{k} e^{\beta' \mathbf{Z}_i(s_j)} \Lambda\{s_j\} 1\{s_j \leq Y_i\}\right)\right)\right). \end{aligned}$$

2. Setting $(\partial l_n(\theta)/\partial \Lambda\{s_j\})|_{\theta = \widehat{\theta}_n} = 0$ and solving for $\Lambda\{s_j\}$.

Solving these two steps, the result is obtained, which concludes the proof.

□

6.6 Consistency

Since we are interested in almost sure (a.s.) consistency, we work with fixed realizations of the data which are assumed to lie in a set of probability one. Let $\|\cdot\|_\infty$ denote the supremum norm on $[0, \tau]$, and recall that $\|\cdot\|$ denotes the Euclidean norm.

Theorem 6.6.1 *Under conditions* ***C1-C7****, the NPMLE is consistent that is,*

$$\sup_{t \in [0,\tau]} |\hat{\Lambda}_n(t) - \Lambda_0(t)|, \quad \|\hat{\gamma}_n - \hat{\gamma}_0\| \quad \textit{and} \quad \|\hat{\beta}_n - \hat{\beta}_0\|$$

converge to 0 almost surely as n *tends to* ∞.

The consistency proof follows the lines of Murphy's proof of a.s. consistency in the frailty model (1994) (see also Zeng and Lin (2007) and Zeng *et al.* (2008), who use similar techniques in transformation models with recurrent events and clustered failure times respectively). However, the technical details are different. Three technical lemmas are needed before presenting the proof.

The following lemma is satisfied, under conditions **C4** and **C5**.

Lemma 6.6.1 *The following inequality holds with probability 1:*

$$\begin{aligned}
&\left\{\pi(\gamma' \boldsymbol{X}_i) e^{\beta' \boldsymbol{Z}_i(Y_i)} \Psi^{(1)}\left(\int_0^{Y_i} e^{\beta' \boldsymbol{Z}_i(s)} d\Lambda(s)\right)\right\}^{\Delta_i} \\
&\qquad \times \left\{1 - \pi(\gamma' \boldsymbol{X}_i) + \pi(\gamma' \boldsymbol{X}_i)\left(1 - \Psi\left(\int_0^{Y_i} e^{\beta' \boldsymbol{Z}_i(s)} d\Lambda(s)\right)\right)\right\}^{(1-\Delta_i)} \\
&\qquad \leq M_4 \left(1 + \Lambda(Y_i)\right)^{-(1+\rho_0)\Delta_i} + M_5 \left(1 + \Lambda(Y_i)\right)^{-(\rho_0 + \Delta_i)}, \qquad (6.11)
\end{aligned}$$

where M_4 and M_5 are positive constants independent of β, γ and Λ.

Proof. By condition **C5**, there exists a constant $\rho_0 > 0$ such that,

$$\limsup_{x \to \infty} (1+x)^{\rho_0}(1 - \Psi(x)) < \infty \qquad \text{and} \qquad \limsup_{x \to \infty} (1+x)^{1+\rho_0}(\Psi^{(1)}(x)) < \infty.$$

By the second inequality, there exists a constant $m_0 < \infty$ such that

$$\pi(\gamma' \mathbf{X}_i) \Psi^{(1)}\left(\int_0^{Y_i} e^{\beta' \mathbf{Z}_i(s)} d\Lambda(s)\right) \leq m_0 \left(1 + \int_0^{Y_i} e^{\beta' \mathbf{Z}_i(s)} d\Lambda(s)\right)^{-(1+\rho_0)}.$$

Hence,

$$\left\{\pi(\gamma' \mathbf{X}_i) \Psi^{(1)}\left(\int_0^{Y_i} e^{\beta' \mathbf{Z}_i(s)} d\Lambda(s)\right)\right\}^{\Delta_i} \leq m_0^{\Delta_i} \left(1 + \int_0^{Y_i} e^{\beta' \mathbf{Z}_i(s)} d\Lambda(s)\right)^{-(1+\rho_0)\Delta_i} \qquad (6.12)$$

By the first inequality of condition **C5**, there exists a constant $m_1 < \infty$ such that

$$\begin{aligned}
1 - \pi(\gamma' \mathbf{X}_i) + \pi(\gamma' \mathbf{X}_i) &\left(1 - \Psi\left(\textstyle\int_0^{Y_i} e^{\beta' \mathbf{Z}_i(s)} d\Lambda(s)\right)\right) \\
&\leq 1 + \left(1 - \Psi\left(\textstyle\int_0^{Y_i} e^{\beta' \mathbf{Z}_i(s)} d\Lambda(s)\right)\right) \\
&\leq 1 + m_1 \left(1 + \textstyle\int_0^{Y_i} e^{\beta' \mathbf{Z}_i(s)} d\Lambda(s)\right)^{-\rho_0},
\end{aligned}$$

and therefore

$$\begin{aligned}\left\{1-\pi(\gamma'\mathbf{X}_i)+\pi(\gamma'\mathbf{X}_i)\left(1-\Psi\left(\int_0^{Y_i}e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right)\right)\right\}^{(1-\Delta_i)} \\ \leq \left\{1+m_1\left(1+\int_0^{Y_i}e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right)^{-\rho_0}\right\}^{(1-\Delta_i)} \\ \leq 1+m_1\left(1+\int_0^{Y_i}e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right)^{-\rho_0(1-\Delta_i)}. \qquad (6.13)\end{aligned}$$

Now, let $m_2=\min\{M_2,1\}$, where M_2 was defined in **C2**. Then,

$$1+\int_0^{Y_i}e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\geq 1+M_2\Lambda(Y_i)\geq m_2\left(1+\Lambda(Y_i)\right). \qquad (6.14)$$

From the inequalities (6.12), (6.13), (6.14) and by (**C2**), we obtain

$$\begin{aligned}&\left\{\pi(\gamma'\mathbf{X}_i)e^{\beta'\mathbf{Z}_i(Y_i)}\Psi^{(1)}\left(\int_0^{Y_i}e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right)\right\}^{\Delta_i} \\ &\qquad\times\left\{1-\pi(\gamma'\mathbf{X}_i)+\pi(\gamma'\mathbf{X}_i)\left(1-\Psi\left(\int_0^{Y_i}e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right)\right)\right\}^{(1-\Delta_i)} \\ &\leq M_1^{\Delta_i}m_0^{\Delta_i}\left[m_2\left(1+\Lambda(Y_i)\right)\right]^{-(1+\rho_0)\Delta_i}\left[1+m_1\left(m_2\left(1+\Lambda(Y_i)\right)\right)^{-\rho_0(1-\Delta_i)}\right] \\ &= M_4\left(1+\Lambda(Y_i)\right)^{-(1+\rho_0)\Delta_i}+M_5\left(1+\Lambda(Y_i)\right)^{-(\rho_0+\Delta_i)}\end{aligned}$$

where $M_4=M_1^{\Delta_i}m_0^{\Delta_i}m_2^{-(1+\rho_0)\Delta_i}$ and $M_5=M_1^{\Delta_i}m_0^{\Delta_i}m_2^{-(\rho_0+\Delta_i)}m_1$. This concludes the proof.

□

Lemma 6.6.2 $\limsup_n\hat{\Lambda}_n(\tau)<\infty$ *almost surely.*

Proof: The key arguments for this proof are based on Murphy (1994), Zeng and Lin (2007) and Zeng *et al.* (2008). Consider the step function $\tilde{\Lambda}_n(\cdot)$ defined as:

$$\tilde{\Lambda}_n(t)=\int_0^t\frac{dG_n(u)}{W_n(u;\theta_0)}.$$

Note that $\tilde{\Lambda}_n(t)$ relates to G_n in a similar manner as $\hat{\Lambda}_n(t)$, but with W_n evaluated at θ_0. Let $\tilde{\theta}_n = (\beta_0, \gamma_0, \tilde{\Lambda}_n)$. The proof proceeds by contradiction. Clearly, $0 \leq n^{-1}[\ln L_n(\hat{\theta}_n) - \ln L_n(\tilde{\theta}_n)]$. The right-hand side in this inequality can be bounded from above by Lemma 6.6.1, leading to

$$\begin{aligned}
0 \leq M_6 + \tfrac{1}{n}\textstyle\sum_{i=1}^n \Big\{ & \Delta_i \ln(\hat{\Lambda}_n\{Y_i\}) \\
& -(1+\rho_0)\Delta_i \ln\left(1+\hat{\Lambda}_n(Y_i)\right) - (\rho_0+\Delta_i)\ln\left(1+\hat{\Lambda}_n(Y_i)\right)\Big\} \\
-\tfrac{1}{n}\textstyle\sum_{i=1}^n \Big\{ & \Delta_i \ln(\tilde{\Lambda}_n\{Y_i\}) + \Delta_i \ln\left(\Psi^{(1)}\left(\textstyle\sum_{j=1}^k e^{\beta_0'\mathbf{Z}_i(s_j)}\tilde{\Lambda}_n\{s_j\}1\{s_j \leq Y_i\}\right)\right) \\
& +(1-\Delta_i)\ln\left(1-\pi(\gamma_0'\mathbf{X}_i)\right. \\
& \left.+\pi(\gamma_0'\mathbf{X}_i)\left(1-\Psi\left(\textstyle\sum_{j=1}^k e^{\beta_0'\mathbf{Z}_i(s_j)}\tilde{\Lambda}_n\{s_j\}1\{s_j \leq Y_i\}\right)\right)\right)\Big\},
\end{aligned}$$

where M_6 is some positive constant. From the construction of $\tilde{\Lambda}_n(t)$, it follows that

$$\begin{aligned}
0 \;&\leq\; M_6 + \frac{1}{n}\sum_{i=1}^n \Big\{\Delta_i \ln(\hat{\Lambda}_n\{Y_i\}) - (2\Delta_i+\rho_0+\rho_0\Delta_i)\ln\left(1+\hat{\Lambda}_n(Y_i)\right)\Big\} \\
&\leq\; M_6 + \frac{1}{n}\sum_{i=1}^n \Big\{\Delta_i \ln(1+\hat{\Lambda}_n(Y_i)) - (2\Delta_i+\rho_0+\rho_0\Delta_i)1\{Y_i=\tau\}\ln\left(1+\hat{\Lambda}_n(Y_i)\right) \\
&\qquad -(2\Delta_i+\rho_0+\rho_0\Delta_i)1\{Y_i<\tau\}\ln\left(1+\hat{\Lambda}_n(Y_i)\right)\Big\}. \qquad (6.15)
\end{aligned}$$

Now, mimicking the arguments in Murphy (1994), we show that the right-hand side of (6.15) is eventually negative if $\hat{\Lambda}_n(\tau)$ diverges. Consider a partition $\tau = s_0 > \ldots > s_N = 0$ of $[0,\tau]$. Then the right-hand side of (6.15) can be bounded from above by

$$\begin{aligned}
M_6 \;&-\; \frac{1}{2n}\sum_{i=1}^n (2\Delta_i+\rho_0+\rho_0\Delta_i)1\{Y_i=\tau\}\ln\left(1+\hat{\Lambda}_n(\tau)\right) \\
&-\; \Big\{\frac{1}{2n}\sum_{i=1}^n (2\Delta_i+\rho_0+\rho_0\Delta_i)1\{Y_i=\tau\}\ln\left(1+\hat{\Lambda}_n(\tau)\right) \\
&\qquad -\frac{1}{n}\sum_{i=1}^n \Delta_i 1\{Y_i\in[s_1,s_0)\}\ln\left(1+\hat{\Lambda}_n(\tau)\right)\Big\} \\
&-\; \Big\{\sum_{q=1}^N \Big\{\frac{1}{n}\sum_{i=1}^n (2\Delta_i+\rho_0+\rho_0\Delta_i)1\{Y_i\in[s_q,s_{q-1})\}\ln\left(1+\hat{\Lambda}_n(s_q)\right) \\
&\qquad -\frac{1}{n}\sum_{i=1}^n \Delta_i 1\{Y_i\in[s_{q+1},s_q)\}\ln\left(1+\hat{\Lambda}_n(s_q)\right)\Big\}\Big\}. \qquad (6.16)
\end{aligned}$$

Using Murphy's idea (1994) for constructing the partition, the sequence $s_0 > s_1 > \ldots > s_N$ can be chosen in such way that the first term in (6.16) diverges to $-\infty$ as $\hat{\Lambda}_n(\tau) \to \infty$ and

the second and third terms are negative for large n. This contradicts the fact that (6.16) should be non-negative. Thus, we have shown that $\limsup \hat{\Lambda}_n(\tau) < \infty$.

□

Lemma 6.6.3 *$\tilde{\Lambda}_n(t)$ converges uniformly to $\Lambda_0(t)$ almost surely.*

Proof. We first show that the class of functions

$$\{w(u, \mathbf{O}; \theta); u \in [0, \tau], \theta \in \Theta\}$$

is a Donsker class. Using the Lemma 2 in Parner (1998), states that if $\mathbf{Z}$ is a caglad process on $[0, \tau]$ which is uniformly bounded in variation then, $\mathbf{Z}(\cdot)$ is Donsker, so by multiplying two Donsker classes we get that $\{\beta'\mathbf{Z}(u) : u \in [0, \tau], \beta \in \mathcal{B}\}$ is Donsker. The exponential function is Lipschitz on compact sets of the real line. (This follows from a first-order Taylor expansion). Since $\mathbf{Z}(\cdot)$ is uniformly bounded, we get from van der Vaart and Wellner (1996), Theorem 2.10.6, that the class $\{e^{\beta'\mathbf{Z}(u)} : u \in [0, \tau], \beta \in \mathcal{B}\}$ is Donsker.

Boundedness of $\mathbf{X}$ and the Theorem 2.7.1 of van der Vaart and Wellner (1996) imply that the class $\{g_\gamma(\mathbf{X}) = \gamma'\mathbf{X} : \gamma \in \mathcal{G}\}$ is Donsker. Differentiability of $e^{\gamma'\mathbf{X}}$ in $\mathbf{X}$ and the boundedness of the derivative imply that $\{e^{\gamma'\mathbf{X}} : \gamma \in \mathcal{G}\}$ is Donsker. By the example 2.10.9 from van der Vaart and Wellner (1996) and the fact that $1+e^{\gamma'\mathbf{X}} > 0$, then $\{e^{\gamma'\mathbf{X}}/(1+e^{\gamma'\mathbf{X}})\}$ is Donsker. By the condition **C4**, the function $\Psi(\cdot)$ is Donsker (its derivative is bounded, then the function is Lipschitz), therefore $1 - \Psi(\cdot)$ is Donsker. By the example 2.10.7 and 2.10.8 of van der Vaart and Wellner (1996) (multiplying and adding classes uniformly bounded Donsker, preserves Donsker property) then the classes, $\{1-\pi(\gamma' X)+\pi(\gamma'\mathbf{X})\Psi(\cdot)\}$ and $\{(1 - \Delta_i)\pi(\gamma'\mathbf{X})\Psi^{(1)}(\cdot)\}$ are Donsker.

As $1 - \pi(\gamma'\mathbf{X}) + \pi(\gamma'\mathbf{X})\Psi(\cdot) > 0$ and by the examples 2.10.8 and 2.10.9 de van der Vaart and Wellner, the class

$$\left\{ \frac{(1 - \Delta)\pi(\gamma'\mathbf{X})\Psi^{(1)}(\cdot)}{1 - \pi(\gamma'\mathbf{X}) + \pi(\gamma'\mathbf{X})\Psi(\cdot)} \right\},$$

is Donsker. By **C4** the functions $\Psi^{(1)}(\cdot)$ and $\Psi^{(2)}(\cdot)$ are Lipschitz with $\Psi^{(1)}(\cdot) > 0$, then,

$$\left\{ \frac{\Delta_i \Psi^{(2)}(\cdot)}{\Psi^{(1)}(\cdot)} \right\}$$

is Donsker. From this analysis, we obtain that $\phi\,(Y, \mathbf{O}; \theta)$ is a Donsker classe. Finally, the indicator function is Donsker.

Therefore, we can conclude that the class

$$\{w(u, \mathbf{O}; \theta); u \in [0, \tau], \theta \in \Theta\}$$

is Donsker. Similar arguments yield that $\{\Delta 1\{u \leq t\}/P_{\theta_0}[w(u, \mathbf{O}; \theta_0)]\}$ is also Donsker class.

Then,

$$\begin{aligned}
\sup_{t\in[0,\tau]} \left|\tilde{\Lambda}_n(t) - \Lambda_0(t)\right| &= \sup_{t\in[0,\tau]} \left| \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i 1\{Y_i \leq t\}}{W_n(Y_i;\theta_0)} - P_{\theta_0}\left[\frac{\Delta 1\{Y \leq t\}}{P_{\theta_0}[w(u,\mathbf{O};\theta_0)]|_{s=Y}}\right]\right| \\
&\leq \sup_{s\in[0,\tau]} \left| \frac{1}{W_n(s;\theta_0)} - \frac{1}{P_{\theta_0}[w(s,\mathbf{O};\theta_0)]}\right| \\
&\quad + \sup_{t\in[0,\tau]} \left| (\mathbb{P}_n - P_{\theta_0}) \left[\frac{\Delta 1\{Y \leq t\}}{P_{\theta_0}[w(u,\mathbf{O};\theta_0)]|_{s=Y}}\right]\right| \qquad (6.17)
\end{aligned}$$

From the result above, $\{w(u, \mathbf{O}; \theta_0); u \in [0, \tau], \theta \in \Theta\}$ is a Donsker class and therefore a Glivenko Cantelli class of functions, and thus $\sup_{u\in[0,\tau]} |W_n(u;\theta_0) - P_{\theta_0}[w(u, \mathbf{O}; \theta_0))]|$ converges to 0 almost surely. Moreover, for $u \in [0, \tau]$, $P_{\theta_0}[w(u, \mathbf{O}; \theta_0)] > 0$ on $[0, \tau]$. Therefore, the first term on the right hand side of (6.17) converges to 0 almost surely and the second term converges almost surely to 0 by the Glivenko-Cantelli property of

$$\{\Delta 1\{u \leq t\}/P_{\theta_0}[w(u, \mathbf{O}; \theta_0)]|_{u=Y}\}.$$

Therefore, we conclude that $\widetilde{\Lambda}_n$ converges uniformly to Λ_0, which concludes the proof.

□

Proof of Theorem 6.6.1. The proof consists of two steps:

1. We show that every subsequence of n contains a further subsequence where the NPMLE $\widehat{\theta}_n$ converges,

2. We show that the set of limits of all convergent subsequences of $\widehat{\theta}_n$ reduces to $\{\theta_0\}$.

Proof of 1. From the compactness of $\mathcal{B} \times \mathcal{G}$ and Bolzanno-Weiersrass's theorem, every subsequence $(\widehat{\beta}_{\phi(n)}, \widehat{\gamma}_{\phi(n)})$of $(\widehat{\beta}_n, \widehat{\gamma}_n)$ has a further subsequence $(\widehat{\beta}_{\varphi(\phi(n))}, \widehat{\gamma}_{\varphi(\phi(n))})$, which converges to some (β^*, γ^*) in $\mathcal{B} \times \mathcal{G}$. By Lemma 6.6.2 and Helly's theorem, we can find with probability 1 a subsequence $\widehat{\Lambda}_{\eta(\varphi(\phi(n)))}$ of $\widehat{\Lambda}_{\varphi(\phi(n))}$ and a nondecreasing right-continuous function Λ^* such that $\widehat{\Lambda}_{\eta(\varphi(\phi(n)))}(t) \to \Lambda^*(t)$ for all $t \in [0, \tau]$ where Λ^* is continuous; $\widehat{\Lambda}_{\eta(\varphi(\phi(n)))}$ is said to converge weakly to Λ^*. In the following, we shall use the following notation for the sake of clarity of formulas $\xi(n) = \eta(\varphi(\phi(n)))$. We now show that Λ^* is

continuous on $[0,\tau]$. Note first that

$$\begin{aligned}\hat{\Lambda}_{\xi(n)}(t) &= \int_0^t \frac{dG_{\xi(n)}(u)}{\mathbb{P}_{\xi(n)}[w(u,\mathbf{O};\hat{\theta}_{\xi(n)})]} \\ &= \int_0^t \frac{\mathbb{P}_{\xi(n)}[w(u,\mathbf{O};\theta_0)]}{\mathbb{P}_{\xi(n)}[w(u,\mathbf{O};\hat{\theta}_{\xi(n)})]} \frac{dG_{\xi(n)}}{\mathbb{P}_{\xi(n)}[w(u,\mathbf{O};\theta_0)]} \\ &= \int_0^t \frac{\mathbb{P}_{\xi(n)}[w(u,\mathbf{O};\theta_0)]}{\mathbb{P}_{\xi(n)}[w(u,\mathbf{O};\theta_{\xi(n)})]} d\tilde{\Lambda}_{\xi(n)}(u), \end{aligned} \tag{6.18}$$

where $\widetilde{\Lambda}_n$ is defined in Lemma 6.6.2.

It follows from the Glivenko-Cantelli property of $\{w(u,\mathbf{O};\theta)\}$ that

$$\begin{aligned}&\sup_{u\in[0,\tau]} \left|\mathbb{P}_{\xi(n)}[w(u,\mathbf{O};\theta_0)] - P_{\theta_0}[w(s,\mathbf{O};\theta_0)]\right| \longrightarrow 0 \quad a.s., \\ &\sup_{u\in[0,\tau]} \left|\mathbb{P}_{\xi(n)}[w(s,\mathbf{O};\widehat{\theta}_{\xi(n)})] - P_{\theta_0}[w(s,\mathbf{O};\widehat{\theta}_{\xi(n)})]\right| \longrightarrow 0 \quad a.s.. \end{aligned} \tag{6.19}$$

Then, $P_0[w(u,\mathbf{O};\hat{\theta}_{\xi(n)})]$ converge uniformly to $P_0[w(u,\mathbf{O};\theta^*)]$. Using this result, the equation 6.19 and the triangle inequality, we obtained that,

$$\frac{d\hat{\Lambda}_{\xi(n)}(t)}{d\tilde{\Lambda}_{\xi(n)}(t)} \to \frac{P_{\theta_0}\left[w(s,\mathbf{0};\theta_0)\right]}{P_{\theta_0}\left[w(s,\mathbf{0};\theta^*)\right]}$$

uniformly in $t \in [0,\tau]$. By taking the limits on both side in (6.18), we obtain

$$\Lambda^*(t) = \int_0^t \frac{P_0\left[w(u,\mathbf{O};\theta_0)\right]}{P_0\left[w(u,\mathbf{O};\theta^*)\right]} d\Lambda_0(u).$$

Thus $\Lambda^*(t)$ is absolutely continuous with respect to $\Lambda_0(t)$, so that $\Lambda^*(t)$ is differentiable with respect to t.

Proof of 2. Finally, we can show that $\beta^* = \beta_0$, $\gamma^* = \gamma_0$, and $\Lambda^* = \Lambda_0$. Consider the difference

$$0 \leq \frac{1}{\xi(n)} l_{\xi(n)}\left(\hat{\gamma}_{\xi(n)}, \hat{\beta}_{\xi(n)}, \hat{\Lambda}_{\xi(n)}\right) - \frac{1}{\xi(n)} l_{\xi(n)}\left(\gamma_0, \beta_0, \tilde{\Lambda}_{\xi(n)}\right).$$

By letting n go to infinity, we obtain that $P_{\theta_0}\left[l\left(\gamma^*,\beta^*,\Lambda^*\right) - l\left(\gamma_0,\beta_0,\Lambda_0\right)\right] \geq 0$. The left-hand side of this inequality is the negative Kullback-Leibler information between the density indexed by θ^* and the true density which implies that $\theta^* = \theta_0$. Thus it has been proven that $\hat{\beta}_n, \hat{\gamma}_n$ and $\hat{\Lambda}(t)$ $(t \in [0,\tau])$ converge almost surely to almost surely β_0, γ_0 and $\Lambda_0(t)$.

□

6.7 Asymptotic normality

6.7.1 Score and Information

To obtain the asymptotic normality, we adapt the function analytic approach developed by Murphy (1995) for the frailty model; see also Fang *et al.* (2005), Kosorok and Song (2007), Lu (2008), and Hernández Quintero *et al.* (2009), who recently adapted this approach to various other semiparametric regression models for survival data. To derive the asymptotic distribution of the estimators, we must verify that an analog of the information matrix (now an operator) is continuously invertible and that the score equations are asymptotically normal. In the latter verification we use results from empirical process theory.

Consider the submodel

$$\epsilon \to \hat{\theta}_{n,\epsilon} = \left(\hat{\beta}_n + \epsilon \mathbf{h}_\beta, \hat{\gamma}_n + \epsilon \mathbf{h}_\gamma, \int_0^{\cdot} (1 + \epsilon h_\Lambda(s))\, d\hat{\Lambda}_n(s) \right)$$

where h_Λ is a non-negative function on $[0, \tau]$, $\mathbf{h}_\beta$ and $\mathbf{h}_\gamma$ are vectors in $\mathbb{R}^q$ and $\mathbb{R}^p$ respectively. Let $\mathbf{h} = (\mathbf{h}_\beta, \mathbf{h}_\gamma, h_\Lambda)$.

Because the maximum likelihood estimator $\hat{\theta}_n = (\hat{\beta}_n, \hat{\gamma}_n, \hat{\Lambda}_n)$ for the full model also maximizes the likelihood under any parametric submodel that passes through $\hat{\theta}_n$, it must satisfy the score function which is obtained by differentiating $l_n(\widehat{\theta}_{n,\epsilon})$ with respect to ϵ, and evaluating at $\epsilon = 0$. That is,

$$S_n(\hat{\theta}_n)(\mathbf{h}) = \left. \frac{\partial l_n(\widehat{\theta}_{n,\epsilon})}{\partial \epsilon} \right|_{\epsilon=0} = 0 \tag{6.20}$$

for every $\mathbf{h}$. Define

$$\begin{aligned}
S_\gamma(\theta) &= \Delta \mathbf{X} \left(1 - \pi(\gamma' \mathbf{X})\right) \\
&\quad - \frac{(1-\Delta)\mathbf{X}\pi(\gamma'\mathbf{X})(1-\pi(\gamma'\mathbf{X}))\Psi\left(\int_0^Y e^{\beta' \mathbf{Z}(s)} d\Lambda(s)\right)}{1 - \pi(\gamma'\mathbf{X}) + \pi(\gamma'\mathbf{X})\left(1 - \Psi\left(\int_0^Y e^{\beta'\mathbf{Z}(s)} d\Lambda(s)\right)\right)}, \\
S_\beta(\theta) &= \Delta \mathbf{Z}(Y) - \phi(\mathbf{O}; \theta) \int_0^Y e^{\beta' \mathbf{Z}(u)} \mathbf{Z}(u) d\Lambda(u), \\
S_\Lambda(\theta)(h_\Lambda) &= \Delta h_\Lambda(Y) - \phi(\mathbf{O}; \theta) \int_0^Y e^{\beta' \mathbf{Z}(u)} h_\Lambda(u) d\Lambda(u).
\end{aligned}$$

Then, the score operator $S_n(\hat{\theta}_n)(\mathbf{h})$ has the form

$$S_n(\hat{\theta}_n)(\mathbf{h}) = \mathbb{P}_n \left[\mathbf{h}'_\beta S_\beta(\hat{\theta}_n) + \mathbf{h}'_\gamma S_\gamma(\hat{\theta}_n) + S_\Lambda(\hat{\theta}_n)(h_\Lambda) \right]. \tag{6.21}$$

We take the space of elements $\mathbf{h}$ to be

$$\mathcal{H} = \left\{ \mathbf{h} = (\mathbf{h}_\beta, \mathbf{h}_\gamma, h_\Lambda) : \mathbf{h}_\beta \in \mathbb{R}^q; \mathbf{h}_\gamma \in \mathbb{R}^p; h_{\Lambda_j} \in VB([0,\tau]) \right\},$$

where VB denotes the bounded variation on $[0,\tau]$. Furthermore, we take the functions h_Λ to be continuous from the right at 0. We define the following norm on $\mathcal{H}$: if $\mathbf{h} \in \mathcal{H}$, let

$$\|\mathbf{h}\|_H = \|\mathbf{h}_\beta\| + \|\mathbf{h}_\gamma\| + \|h_\Lambda\|_v,$$

where $\|\cdot\|$ is the Euclidean norm and $\|h_\Lambda\|_v$ denotes the total variation of h_Λ on $[0,\tau]$. We further define $\mathcal{H}_r = \{\mathbf{h} \in \mathcal{H}, \|\mathbf{h}\|_\mathcal{H} \leq r\}$ and $\mathcal{H}_\infty = \{\mathbf{h} \in \mathcal{H}, \|\mathbf{h}\|_\mathcal{H} < \infty\}$.

Define

$$\theta(\mathbf{h}) = \mathbf{h}'_\beta \beta + \mathbf{h}'_\gamma \gamma + \int_0^\tau h_\Lambda(s) d\Lambda(s),$$

where $\mathbf{h} \in \mathcal{H}$. From this, we can re-consider the parameter θ as a linear functional on $\mathcal{H}_r$, and the parameter space Θ as a subset of $l^\infty(\mathcal{H}_r)$ which is the space of bounded real-valued functions on $\mathcal{H}_r$. Moreover, the score operator S_n appears to be a random map from Θ to the space $l^\infty(\mathcal{H}_r)$.

Remark. Note that appropriate choices for $\mathbf{h}$ allow to extract all components of the original parameter θ; in the present study, we shall denote by $\mathbf{0}_r$ ($r \geq 2$) the r-dimensional column vector having all its components equal to 0.

For example, let $\mathbf{h}_\gamma = \mathbf{0}_p$, $h_\Lambda(\cdot) = 0$, and $\mathbf{h}_\beta$ be the q-dimensional vector with a one at the i-th location and zeros elsewhere. This yields the i-th component of β.

As an another example, let $\mathbf{h}_\beta = \mathbf{0}_q$, $\mathbf{h}_\gamma = \mathbf{0}_p$ and $h_\Lambda(\cdot) = 1\{\cdot \leq t\}$, for some $t \in (0,\tau)$. In this case, $\theta(\mathbf{h})$ reduces to $\Lambda(t)$.

Similar to parametric model, we define the Fisher information operator by

$$I(\theta_0)(\mathbf{h}) = P_{\theta_0}[S_1(\theta_0)(\mathbf{h})^2]$$

where S_1 is the score operator (6.21) based on a single observation. An explicit expression of $I(\theta_0)(\mathbf{h})$ is given in the lemma below.

Lemma 6.7.1 *Let $\mathbf{h} \in \mathcal{H}_r$. Then $P_{\theta_0}[S_1(\theta_0)(\mathbf{h})] = 0$ and the Fisher information operator is given by*

$$I(\theta_0)(\mathbf{h}) = \mathbf{h}'_\beta \sigma_\beta(\mathbf{h}) + \mathbf{h}'_\gamma \sigma_\gamma(\mathbf{h}) + \int_0^\tau \sigma_\Lambda(\mathbf{h})(s) h_\Lambda(s) d\Lambda_0(s), \tag{6.22}$$

where

$$\begin{aligned}
\sigma_\beta(\mathbf{h}) &= P_{\theta_0}\left[2S_\beta(\theta_0)\Delta h_\Lambda(Y)\right] + P_{\theta_0}\left[S_\beta(\theta_0)^{\otimes 2}\right]\mathbf{h}_\beta + P_{\theta_0}\left[S_\beta(\theta_0)S_\gamma(\theta_0)'\right]\mathbf{h}_\gamma \\
\sigma_\gamma(\mathbf{h}) &= P_{\theta_0}\left[2S_\gamma(\theta_0)\Delta h_\Lambda(Y)\right] + P_{\theta_0}\left[S_\gamma(\theta_0)^{\otimes 2}\right]\mathbf{h}_\gamma + P_{\theta_0}\left[S_\gamma(\theta_0)S_\beta(\theta_0)'\right]\mathbf{h}_\beta \\
\sigma_\Lambda(\mathbf{h})(s) &= -P_{\theta_0}\left[\Delta h_\Lambda(Y)1\{s\le Y\}\phi(\mathbf{O};\theta_0)e^{\beta_0'\mathbf{Z}(s)}\right] \\
&\quad + P_{\theta_0}\left[\{\phi(\mathbf{O};\theta_0)\}^2\left\{\int_0^Y e^{\beta_0'\mathbf{Z}(u)}h_\Lambda(u)d\Lambda_0(u)\right\}1\{s\le Y\}e^{\beta_0'\mathbf{Z}(s)}\right] \\
&\quad - \mathbf{h}_\beta' P_{\theta_0}\left[2S_\beta(\theta_0)1\{s\le Y\}\phi(\mathbf{O};\theta_0)e^{\beta_0'\mathbf{Z}(s)}\right] \\
&\quad - \mathbf{h}_\gamma' P_{\theta_0}\left[2S_\gamma(\theta_0)1\{s\le Y\}\phi(\mathbf{O};\theta_0)e^{\beta_0'\mathbf{Z}(s)}\right],
\end{aligned}$$

where for any r-dimensional vector u, $u^{\otimes 2} = u'u$.

Proof: We first prove that $P_{\theta_0}\left[S_1(\theta_0)(\mathbf{h})\right] = 0$. Note that

$$\begin{aligned}
P_{\theta_0}\left[S_\beta(\theta_0)\right] &= P_{\theta_0}\left[\Delta\mathbf{Z}(Y) - \phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}\mathbf{Z}(u)d\Lambda(u)\right] \\
&= P_{\theta_0}\left[\int_0^\tau \mathbf{Z}(u)dN(u) - \phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}\mathbf{Z}(u)d\Lambda(u)\right] \\
&= P_{\theta_0}\left[\int_0^\tau \mathbf{Z}(u)dM(u)\right]
\end{aligned}$$

where,

$$M(t) = N(t) - \int_0^t 1\{Y\ge u\}\phi(\mathbf{O};\theta)e^{\beta'\mathbf{Z}(u)}d\Lambda(u)$$

is a counting process martingale with respect to the filtration $\sigma\{N(s), 1\{Y\le s, \Delta = 1\}, \mathbf{X}, \mathbf{Z}(s) : 0\le s\le t\}$. Note that, we have obtained a process in Y, which is a martingale stochastic integral, provided the time-dependent covariate, which is predictable (to verify this is enough to see that is bounded, and this is obtained by **C2**). Therefore $P_{\theta_0}\left[S_\beta(\theta_0)\right] = 0$.

Similar arguments imply that $P_{\theta_0}\left[S_\Lambda(\theta_0)\right] = 0$ and $P_{\theta_0}\left[S_\gamma(\theta_0)\right] = 0$. This conclude the firs part of the proof.

To prove the second result, we develop

$$S_1(\theta_0)(\mathbf{h})^2 = \left[\mathbf{h}_\beta' S_\beta(\theta_0) + \mathbf{h}_\gamma' S_\gamma(\theta_0) + S_\Lambda(\theta_0)(h_\Lambda)\right]^2,$$

and we take the expectation of the resulting expression. By the first part, we have that $P_{\theta_0}\left[S_\Lambda(\theta_0)(h_\Lambda)\right] = 0$ for any bounded function h_Λ, which implies that

$$P_{\theta_0}\left[\Delta h_\Lambda(Y)\right] = P_{\theta_0}\left[\phi(\mathbf{O};\theta)\int_0^Y e^{\beta_0'\mathbf{Z}(u)}h_\Lambda(u)d\Lambda_0(u)\right].$$

Some lengthy algebraic manipulations and re-arrangement of terms yield the result.

□

Lemma 6.7.2 *The operator* $\sigma = (\sigma_\beta, \sigma_\gamma, \sigma_\Lambda) : \mathcal{H}_r \to \mathcal{H}_r$ *is one-to-one.*

Proof: Assume $\sigma(\mathbf{h}) = 0$. By Lemma 6.7.1, $P_{\theta_0}\left[S_1(\theta_0)(\mathbf{h})^2\right] = 0$. It follows that $S_1(\theta_0)(\mathbf{h}) = 0$ almost surely.

We successively take $\Delta = 1$ and $\Delta = 0$. We obtain 2 equations. Then, we take the difference between them and some algebraic manipulation and re-arrangement of terms yield the following equation:

$$\mathbf{h}'_\gamma \left[\mathbf{X}\left(1 - \pi(\gamma_0'\mathbf{X})\right) + \frac{\mathbf{X}\pi(\gamma_0'\mathbf{X})(1-\pi(\gamma_0'\mathbf{X}))\Psi\left(\int_0^Y e^{\beta_0'\mathbf{Z}(s)} d\Lambda_0(s)\right)}{1 - \pi(\gamma_0'\mathbf{X}) + \pi(\gamma_0'\mathbf{X})\left(1 - \Psi\left(\int_0^Y e^{\beta_0'\mathbf{Z}(s)} d\Lambda_0(s)\right)\right)} \right]$$
$$+\mathbf{h}'_\beta \mathbf{Z}(Y) + h_\Lambda(Y) + \Gamma(\theta_0) \int_0^Y e^{\beta_0'\mathbf{Z}(s)} \left\{\mathbf{h}'_\beta \mathbf{Z}(s) + h_\Lambda(s)\right\} d\Lambda_0(s) = 0, \qquad (6.23)$$

where $\Gamma(\theta_0) = \left(\frac{\pi(\gamma_0'\mathbf{X})\Psi^{(1)}\left(\int_0^Y e^{\beta_0'\mathbf{Z}(s)} d\Lambda_0(s)\right)}{1-\pi(\gamma_0'\mathbf{X})+\pi(\gamma_0'\mathbf{X})\left(1-\Psi\left(\int_0^Y e^{\beta_0'\mathbf{Z}(s)} d\Lambda_0(s)\right)\right)} + \frac{\Psi^{(2)}\left(\int_0^Y e^{\beta_0'\mathbf{Z}(s)} d\Lambda_0(s)\right)}{\Psi^{(1)}\left(\int_0^Y e^{\beta_0'\mathbf{Z}(s)} d\Lambda_0(s)\right)} \right).$

By choosing Y arbitrarily close to 0, (6.23) reduces to

$$\mathbf{h}'_\gamma \mathbf{X}\left(1 - \pi(\gamma_0'\mathbf{X})\right) + \mathbf{h}'_\beta \mathbf{Z}(0) + h_\Lambda(0) = 0,$$

since h_Λ and Λ_0 are continuous from the right at 0 and $\Lambda_0(0) = 0$, this expression is equivalent,

$$\mathbf{h}'_\gamma \mathbf{X}\left(1 - \pi(\gamma_0'\mathbf{X})\right) = -\mathbf{h}'_\beta \mathbf{Z}(0) - h_\Lambda(0),$$

note that the right-hand side of the previous equation not depend on $\mathbf{X}$ then $\mathbf{h}_\gamma$ must equal 0. So the expression reduces (6.23) to

$$\mathbf{h}'_\beta \mathbf{Z}(Y) + h_\Lambda(Y) + \Gamma(\theta_0) \int_0^Y e^{\beta_0'\mathbf{Z}(s)} \left\{\mathbf{h}'_\beta \mathbf{Z}(s) + h_\Lambda(s)\right\} d\Lambda_0(s) = 0. \qquad (6.24)$$

Note that we have obtained a homogeneous equation for $\mathbf{h}'_\beta \mathbf{Z}(t) + h_\Lambda(t)$ (i.e., there is no isolated constant term in the equation) which has only the trivial solution (see Zeng and Lin, 2007 and technical report of Zeng and Lin, 2006). Therefore, $\mathbf{h}'_\beta \mathbf{Z}(t) + h_\Lambda(t) = 0$ for all $t \in [0, \tau]$. By condition **C7**, it follows that $\mathbf{h}_\beta = 0$ and $h_\Lambda = 0$. This concludes the proof.

□

Lemma 6.7.3 *The operator σ is continuously invertible with an inverse σ^{-1} denoted as $\sigma^{-1} = (\sigma_\beta^{-1}, \sigma_\gamma^{-1}, \sigma_\Lambda^{-1})$.*

Proof: Since $\mathcal{H}_r$ is a Banach space, to prove that σ is continuously invertible, it is sufficient to prove that σ is one-to-one and that it can be written as the sum $A+(\sigma-A)$ of a bounded linear operator A with bounded inverse and a compact operator $\sigma - A$ (Lemma 25.93 of van der Vaart, 1998).

σ is one-to-one by Lemma 6.7.2. Next, define the linear operator $A : \mathcal{H}_r \to \mathcal{H}_r$ by $A(\mathbf{h}) = (\mathbf{h}_\beta, \mathbf{h}_\gamma, h_\Lambda(\cdot)P_{\theta_0}[W(\cdot, \mathbf{O}; \theta_0)])$. A is bounded (by **C1** and **C2**). In addition, $P_{\theta_0}[W(\cdot, \mathbf{O}, \theta_0)]$ is uniformly bounded away from 0 on $[0, \tau]$. Thus A is invertible with bounded inverse $A^{-1}(\mathbf{h}) = (\mathbf{h}_\beta, \mathbf{h}_\gamma, h_\Lambda(\cdot)P_{\theta_0}[W(\cdot, \mathbf{O}, \theta_0)]^{-1})$.

The operator $\sigma - A$ is compact, by using the same techniques as in Lu (2008) we can get the result. Because a bounded linear operator with finite dimensional range is compact, we only need show that the operator $K_\Lambda : VB(0, \tau) \to VB(0, \tau)$, given by

$$K_\Lambda(h_\Lambda)(s) = -P_{\theta_0}\left[\Delta h_\Lambda(Y) 1\{s \le Y\}\phi(\mathbf{O}; \theta_0) e^{\beta_0' \mathbf{Z}(s)}\right]$$

$$+P_{\theta_0}\left[\{\phi(\mathbf{O}; \theta_0)\}^2 \left\{\int_0^Y e^{\beta_0' \mathbf{Z}(u)} h_{\Lambda(u)} d\Lambda_0(u)\right\} 1\{s \le Y\} e^{\beta_0' \mathbf{Z}(s)}\right]$$

is compact.

Thus given a sequence of function $h_{\Lambda,n}$ with $\|h_{\Lambda,n}\|_v \le 1$, we must show that there exits a subsequence and an element $g \in VB(0, \tau)$ such that $\|K_\Lambda h_{\Lambda,\eta(n)} - g\|_v \to 0$.

Now, note that K_Λ is a linear operator then, $\|K_\Lambda h_\Lambda\|_v \le M_7 \int |h_\Lambda(u)| d\Lambda_0(u)$ for every h_Λ and a fixed constant M_7. Hence it suffices to show that there exits a subsequence $h_{\Lambda,\eta(n)}$ of $h_{\Lambda,n}$ that converges. Since h_Λ is of bounded variation, we can write $h_{\Lambda,n}$ as the difference of bounded increasing function $h_{\Lambda,n}^{(1)}$ and $h_{\Lambda,n}^{(2)}$. From Helly's theorem, there exists a subsequence $h_{\Lambda,\eta(n)}^{(1)}$ of $h_{\Lambda,n}^{(1)}$ which converges pointwise to some $h_\Lambda^{(1)*}$. There also exists a subsequence $h_{\Lambda,\eta(n)}^{(2)}$ of $h_{\Lambda,n}^{(2)}$ which converges pointwise to some $h_\Lambda^{(2)*}$. Then $h_{\Lambda,n}$ converge to the difference of the limits by the dominated convergence theorem. It follows that $\sigma - A$ is a compact operator.

□

6.7.2 Asymptotic normality result

Theorem 6.7.1 *Under conditions **C1-C7**,*

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, \Sigma_\beta) \qquad \text{and} \qquad \sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow{d} N(0, \Sigma_\gamma)$$

where

$$\Sigma_\beta = \left(\sigma_\beta^{-1}(\mathbf{e}_1, \mathbf{0}_p, 0), \ldots, \sigma_\beta^{-1}(\mathbf{e}_q, \mathbf{0}_p, 0)\right)$$

and

$$\Sigma_\gamma = \left(\sigma_\gamma^{-1}(\mathbf{0}_q, \mathbf{d}_1, 0), \ldots, \sigma_\gamma^{-1}(\mathbf{0}_q, \mathbf{d}_p, 0)\right)$$

are the efficient variances for estimating β_0 and γ_0, and $\mathbf{e}_i$ is the q-dimensional vector with the i-th component 1 and elsewhere 0, and $\mathbf{d}_i$ the p-dimensional vector with the i-th component 1 and elsewhere 0. Furthermore, for any $t \in [0, \tau]$,

$$\sqrt{n}\left(\hat{\Lambda}_n(t) - \Lambda_0(t)\right) \xrightarrow{d} N(0, \upsilon^2(t)),$$

where $\upsilon^2(t) = \int_0^t \sigma_\Lambda^{-1}\left(\mathbf{0}_q, \mathbf{0}_p, 1\{u \le t\}\right) d\Lambda_0(u)$.

The following lemma is needed to prove Theorem 6.7.1.

Lemma 6.7.4 *The efficient score functions for estimating β_0 and γ_0 are*

$$l_\beta = S_\beta(\theta_0) - S_\Lambda(\theta_0)(g_\beta^*) - \Sigma_{23} S_\gamma(\theta_0) \tag{6.25}$$

and

$$l_\gamma = S_\gamma(\theta_0) - S_\Lambda(\theta_0)(g_\gamma^*) - \Sigma_{32} S_\beta(\theta_0), \tag{6.26}$$

respectively, where

$$\Sigma_{23} = -\Sigma_\beta^{-1} \begin{pmatrix} \sigma_\gamma^{-1}(\mathbf{e}_1, \mathbf{0}_p, 0)' \\ \vdots \\ \sigma_\gamma^{-1}(\mathbf{e}_q, \mathbf{0}_p, 0)' \end{pmatrix} \qquad \text{and} \qquad \Sigma_{32} = -\Sigma_\gamma^{-1} \begin{pmatrix} \sigma_\beta^{-1}(\mathbf{0}_q, \mathbf{d}_1, 0)' \\ \vdots \\ \sigma_\beta^{-1}(\mathbf{0}_q, \mathbf{d}_p, 0)' \end{pmatrix},$$

and S_Λ is applied componentwise to the vectors

$$g_\beta^* = -\Sigma_\beta^{-1} \begin{pmatrix} \sigma_\Lambda^{-1}(\mathbf{e}_1, \mathbf{0}_p, 0)' \\ \vdots \\ \sigma_\Lambda^{-1}(\mathbf{e}_q, \mathbf{0}_p, 0)' \end{pmatrix} \qquad \text{and} \qquad g_\gamma^* = -\Sigma_\gamma^{-1} \begin{pmatrix} \sigma_\Lambda^{-1}(\mathbf{0}_q, \mathbf{d}_1, 0)' \\ \vdots \\ \sigma_\Lambda^{-1}(\mathbf{0}_q, \mathbf{d}_p, 0)' \end{pmatrix}.$$

Moreover, the efficient asymptotic variance matrices of $\hat{\beta}_n$ and $\hat{\gamma}_n$ are

$$(P_{\theta_0}[l_\beta l_\beta])^{-1} = \Sigma_\beta \qquad \textit{and} \qquad (P_{\theta_0}[l_\gamma l_\gamma])^{-1} = \Sigma_\gamma,$$

respectively.

Proof: We must show that l_β is orthogonal to the score $S_\Lambda(\theta_0)(g)$ for any bounded function g. Consider $\mathbf{e}_i'\Sigma_\beta \mathbb{P}_{\theta_0}[l_\beta S_\Lambda(\theta_0)(g)]$, wich is equal to

$$P_{\theta_0}\left[\left(e_i'\Sigma_\beta S_\beta(\theta_0) - S_\Lambda(\theta_0)(e_i'\Sigma_\beta g_\beta^*) - e_i'\Sigma_\beta\Sigma_{23}S_\gamma(\theta_0)\right) S_\Lambda(\theta_0)(g)\right] \tag{6.27}$$

Now $e_i'\Sigma_\beta = \sigma_\beta^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)'$, then the equation (6.27) is equivalently to

$$P_{\theta_0}\left(\left[\sigma_\beta^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)'S_{1\beta}(\theta_0) + S_\Lambda(\theta_0)(\sigma_\Lambda^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)) + \sigma_\gamma^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)'S_\gamma(\theta_0)\right) S_\Lambda(\theta_0)(g)\right]. \tag{6.28}$$

As

$$P_{\theta_0}\left[\Delta h_\Lambda(Y)\right] = P_{\theta_0}\left[\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}h_\Lambda(u)d\Lambda(u)\right]. \tag{6.29}$$

Developing the terms in equation (6.28), we obtained $S_\beta(\theta_0)S_\Lambda(\theta_0)(g)$ is equivalently to

$$\begin{aligned}
&\Delta\mathbf{Z}(Y)\Delta g_\Lambda(Y) - \Delta\mathbf{Z}(Y)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}g_\Lambda(u)d\Lambda(u)\\
&-\Delta g_\Lambda(Y)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}\mathbf{Z}(u)d\Lambda(u)\\
&+\phi^2(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}\mathbf{Z}(u)d\Lambda(u)\int_0^Y e^{\beta'\mathbf{Z}(u)}g_\Lambda(u)d\Lambda(u).
\end{aligned}$$

Using the equation (6.29) above is equivalent to

$$\begin{aligned}
&\Delta\mathbf{Z}(Y)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}g_\Lambda(u)d\Lambda(u) - \Delta\mathbf{Z}(Y)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}g_\Lambda(u)d\Lambda(u)\\
&-\Delta g_\Lambda(Y)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}\mathbf{Z}(u)d\Lambda(u)\\
&+\phi^2(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}\mathbf{Z}(u)d\Lambda(u)\int_0^Y e^{\beta'\mathbf{Z}(u)}g_\Lambda(u)d\Lambda(u)
\end{aligned}$$

reducing terms this equals to $\int g\sigma_\Lambda(\mathbf{e}_i, \mathbf{0}_p, 0)$. Similarly, the term

$$S_\Lambda(\theta_0)(\sigma_\Lambda^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0))S_\Lambda(\theta_0)(g)$$

is equivalently to

$$\begin{aligned}
&\Delta\sigma_\Lambda^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)\Delta g(Y) - \Delta\sigma_\Lambda^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}g(u)d\Lambda(u)\\
&-\Delta g(Y)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}\sigma_\Lambda^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)d\Lambda(u) + \phi^2(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}\sigma_\Lambda^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)d\Lambda(u).
\end{aligned}$$

Using the equation (6.29) above is equivalent to

$$\Delta\sigma_\Lambda^{-1}(\mathbf{e}_i,\mathbf{0}_p,0)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}g(u)d\Lambda(u) - \Delta\sigma_\Lambda^{-1}(\mathbf{e}_i,\mathbf{0}_p,0)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}g(u)d\Lambda(u)$$
$$-\Delta g(Y)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}\sigma_\Lambda^{-1}(\mathbf{e}_i,\mathbf{0}_p,0)d\Lambda(u) + \phi^2(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}\sigma_\Lambda^{-1}(\mathbf{e}_i,\mathbf{0}_p,0)d\Lambda(u)$$

reducing terms this equals to $\int g\sigma_\Lambda(\sigma^{-1}(\mathbf{0}_q,\mathbf{e}_i,0))$. Finally, the term

$$S_\gamma(\theta_0)S_\Lambda(\theta_0)(g)$$

is equal to

$$\Delta\mathbf{X}\left(1-\pi(\gamma'\mathbf{X})\right)\Delta g(Y) - \Delta\mathbf{X}\left(1-\pi(\gamma'\mathbf{X})\right)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}gd\Lambda(u)$$
$$-\frac{\Delta g(Y)(1-\Delta)\mathbf{X}\pi(\gamma'\mathbf{X})(1-\pi(\gamma'\mathbf{X}))\Psi\left(\int_0^Y e^{\beta'\mathbf{Z}(s)}d\Lambda(s)\right)}{1-\pi(\gamma'\mathbf{X})+\pi(\gamma'\mathbf{X})\left(1-\Psi\left(\int_0^Y e^{\beta'\mathbf{Z}(s)}d\Lambda(s)\right)\right)}$$
$$+\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}g(u)d\Lambda(u)\frac{\Delta g(Y)(1-\Delta)\mathbf{X}\pi(\gamma'\mathbf{X})(1-\pi(\gamma'\mathbf{X}))\Psi\left(\int_0^Y e^{\beta'\mathbf{Z}(s)}d\Lambda(s)\right)}{1-\pi(\gamma'\mathbf{X})+\pi(\gamma'\mathbf{X})\left(1-\Psi\left(\int_0^Y e^{\beta'\mathbf{Z}(s)}d\Lambda(s)\right)\right)}.$$

Using the equation (6.29) and reducing the above terms is equivalent to $\int g\sigma_\Lambda(\mathbf{0}_q,\mathbf{e}_i,0)$. From the results obtained is reduced to $\mathbf{e}_i'\Sigma_\beta\mathbb{P}_{\theta_0}[l_\beta S_\Lambda(\theta_0)(g)] = \int g\sigma_\Lambda(\sigma^{-1}(\mathbf{e}_i,\mathbf{0}_p,0))d\Lambda_0 = 0$. It follows from Theorem 3.4.1 of Bickel *et al.* (1993) that l_β is the efficient score function for estimating β.

That l_γ is the efficient score function for estimating γ can be proved along similar lines.

Because

$$\mathbf{e}_i\mathbb{P}_{\theta_0}[\Sigma_\beta l_\beta l_\beta'\Sigma_\beta]\mathbf{e}_j = \mathbb{P}_{\theta_0}[\sigma_\beta^{-1}(\mathbf{e}_j,\mathbf{0}_p,0)'S_\beta(\theta_0)S_\beta'(\theta_0)\sigma_\beta^{-1}(\mathbf{e}_i,\mathbf{0}_p,0)'] = \mathbf{e}_i'\Sigma_\beta\mathbf{e}_j$$

for all $i,j=1,\ldots,q$, the second equality is obtained from the Lemma 6.7.1. Then, we have $\mathbb{P}_{\theta_0}[\Sigma_\beta l_\beta l_\beta'\Sigma_\beta] = \Sigma_\beta$ wich implies that $(\mathbb{P}_{\theta_0}[l_\beta l_\beta'])^{-1} = \Sigma_\beta$. Similarly $(\mathbb{P}_{\theta_0}[l_\gamma l_\gamma'])^{-1} = \Sigma_\gamma$.

□

Proof of Theorem 6.7.1: The proof follows the ideas developed in Theorem 5.7.1. Similarly we can show that

$$\sqrt{n}\left(\mathbf{h}_\beta'\left(\widehat{\beta}_n-\beta_0\right)+\mathbf{h}_\gamma'\left(\widehat{\gamma}_n-\gamma_0\right)+\int_0^\tau h_\Lambda(s)d\left(\widehat{\Lambda}_n-\Lambda_0\right)(s)\right) =$$
$$\sqrt{n}\left(S_n(\theta_0)(\sigma^{-1}(\mathbf{h})) - P_{\theta_0}\left[S_1(\theta_0)(\sigma^{-1}(\mathbf{h}))\right]\right)+o_p(1) \qquad (6.30)$$

uniformly in $\mathbf{h}$ as $n \to \infty$.

Setting $\mathbf{h}_\gamma = \mathbf{0}_p$ and h_Λ be identically equal to 0. The equation (6.30) reduces to

$$\sqrt{n}\mathbf{h}'_\beta \left(\widehat{\beta}_n - \beta_0\right) = \sqrt{n}\left(S_n(\theta_0)(\sigma^{-1}(\mathbf{h}_\beta, \mathbf{0}_p, 0)) - P_{\theta_0}\left[S_1(\theta_0)(\sigma^{-1}(\mathbf{h}_\beta, \mathbf{0}_p, 0))\right]\right) + o_p(1),$$

By the central limit theorem, $\sqrt{n}\mathbf{h}'_\beta(\hat{\beta}_n - \beta_0)$ is asymptotically normal with mean 0 and variance

$$P_{\theta_0}[S(\sigma^{-1}(\mathbf{h}_\beta, \mathbf{0}_p, 0))]^2 = [(\sigma_\beta^{-1}(\mathbf{h}_\beta, \mathbf{0}_p, 0))]'\mathbf{h}_\beta = \mathbf{h}'_\beta \Sigma_\beta \mathbf{h}_\beta, \tag{6.31}$$

for any $\mathbf{h}_\beta$ where the first equality follows from 6.22. Therefore, by Cramer-Wold device (van der Vaart, 1998), $\sqrt{n}(\hat{\beta}_n - \beta_0)$ is asymptotically normal with variance Σ_β. Furthermore, by Lemma 6.7.4, Σ_β is efficient variance.

That $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$ is asymptotically normal with mean 0 and variance Σ_γ is proved similarly by letting $h_\Lambda = 0$ and $\mathbf{h}_\beta = \mathbf{0}_q$ in 6.30.

Finally, plugging $\mathbf{h} = (\mathbf{0}_q, \mathbf{0}_p, 1\{u \leq t\})$ we have

$$\begin{aligned}\sqrt{n}(\widehat{\Lambda}_n(t) - \Lambda_0(t)) = \sqrt{n}\big(&S_n(\theta_0)(\sigma^{-1}(\mathbf{0}_q, \mathbf{0}_p, 1\{u \leq t\})) \\ &- P_{\theta_0}\left[S_1(\theta_0)(\sigma^{-1}(\mathbf{0}_q, \mathbf{0}_p, 1\{u \leq t\}))\right]\big) + o_p(1).\end{aligned}$$

Which has an asymptotic normal distribution with mean 0 and variance

$$\begin{aligned}\upsilon^2(t) = &\quad P_{\theta_0}[S_1(\sigma^{-1}(\mathbf{0}_q, \mathbf{0}_p, 1\{u \leq t\}))]^2 \\ = &\quad I(\theta_0)(\sigma^{-1}(\mathbf{0}_q, \mathbf{0}_p, 1\{u \leq t\})) \\ = &\quad \int_0^t \sigma_\Lambda^{-1}\left(\mathbf{0}_q, \mathbf{0}_p, 1\{u \leq t\}\right) d\Lambda_0(u).\end{aligned}$$

This completes the proof.

□

6.8 Variance estimation

The asymptotic variances of $\hat{\beta}_n$, $\hat{\gamma}_n$ and $\hat{\Lambda}_n$ involve inverting a linear operator σ in a functional space. Because the inverse σ^{-1} has no closed form, estimation of the asymptotic variances is not straightforward. One possible method for estimating the asymptotic variances of the Euclidian regression parameter estimates $\hat{\beta}_n$ and $\hat{\Lambda}_n$ is to invert an observed discrete information matrix, which suggested by Sy and Taylor (2000). Another possible approach is to derive a variation of profile likelihood methods Nielsen *et al.* (1992) and

Murphy *et al.* (1997), which would required numerical differentiation of a profile likelihood. However, it is not clear how this methods can be used to given a consistent estimates for the asymptotic variances. Below we give consistent estimates of the asymptotic variances for both the Euclidian parameter estimates $\hat{\beta}_n$, $\hat{\gamma}_n$ and the infinite-dimensional parameter estimate $\hat{\Lambda}_n$. We follow the approach developed by Fang *et al.* (2005).

Define the $(q \times q)$, $(q \times p)$, $(p \times q)$ and $(q \times q)$ matrices $\mathbb{A}_n^\beta$, $\mathbb{A}_n^\gamma$, $\mathbb{B}_n^\beta$, and $\mathbb{B}_n^\gamma$ by

$$\begin{aligned}
\mathbb{A}_n^\beta &= \mathbb{P}_n \left[S_\beta(\widehat{\theta}_n)^{\otimes 2} \right], \\
\mathbb{B}_n^\gamma &= \mathbb{P}_n \left[S_\gamma(\widehat{\theta}_n)^{\otimes 2} \right], \\
\mathbb{A}_n^\gamma &= \mathbb{P}_n \left[S_\beta(\widehat{\theta}_n) S_\gamma(\widehat{\theta}_n)' \right] = \left(\mathbb{B}_n^\beta \right)'.
\end{aligned}$$

Define the $(q \times k)$ matrix

$$\mathbb{A}_n^\Lambda = \frac{2}{n} S_\beta(\widehat{\theta}_n).$$

Similarly, define the $(Q \times s_n)$ partitioned matrix

$$\mathbb{B}_n^\Lambda = \frac{2}{n} S_\gamma(\widehat{\theta}_n).$$

Let $(k \times q)$ and $(k \times p)$ matrices

$$\begin{aligned}
\mathbb{C}_n^\beta &= -\mathbb{P}_n \left[2 S_\beta(\widehat{\theta}_n)' \phi(Y, \mathbf{O}; \hat{\theta}_n) e^{\hat{\beta}_n' \mathbf{Z}(s)} 1\{s \leq Y\} \right], \\
\mathbb{C}_n^\gamma &= -\mathbb{P}_n \left[2 S_\gamma(\widehat{\theta}_n)' \phi(Y, \mathbf{O}; \hat{\theta}_n) e^{\hat{\beta}_n' \mathbf{Z}(s)} 1\{s \leq Y\} \right],
\end{aligned}$$

Next, let the $(k \times k)$ matrix $\mathbb{C}_n^\Lambda$ defined as follows by its (l, m)-th element:

$$\begin{aligned}
\mathbb{C}_n^\Lambda(l, m) = \; & \mathbb{P}_n \left[\left\{ \phi(Y, \mathbf{O}; \hat{\theta}_n) \right\}^2 \widehat{\Delta\Lambda}(Y_m) e^{\hat{\beta}_n' \mathbf{Z}(Y_m)} 1\{Y_m \leq Y\} \right] \\
& -1\{l = m\} \mathbb{P}_n \left[\phi(Y, \mathbf{O}; \hat{\theta}_n) e^{\hat{\beta}_n' \mathbf{Z}(s)} 1\{Y_m \leq Y\} \right]
\end{aligned}$$

where, $\widehat{\Delta\Lambda}(t)$ denotes the jump size of $\widehat{\Lambda}$ at time t; that is, $\widehat{\Delta\Lambda}(t) = \widehat{\Lambda}(t) - \widehat{\Lambda}(t-)$.

Define the partitioned matrix

$$\mathbb{D}_n = \begin{pmatrix} \mathbb{A}_n^\beta & \mathbb{A}_n^\gamma & \mathbb{A}_n^\Lambda \\ \mathbb{B}_n^\beta & \mathbb{B}_n^\gamma & \mathbb{B}_n^\Lambda \\ \mathbb{C}_n^\beta & \mathbb{C}_n^\gamma & \mathbb{C}_n^\Lambda \end{pmatrix}$$

and the matrices

$$\begin{aligned}
\Sigma_{\beta,n} &= \big\{\mathbb{A}_n^{\beta} - \mathbb{A}_n^{\gamma}(\mathbb{B}_n^{\gamma})^{-1}\mathbb{B}_n^{\beta} - (\mathbb{A}_n^{\Lambda} - \mathbb{A}_n^{\gamma}(\mathbb{B}_n^{\gamma})^{-1}\mathbb{B}_n^{\Lambda}) \\
&\qquad \times(\mathbb{C}_n^{\Lambda} - \mathbb{C}_n^{\gamma}(\mathbb{B}_n^{\gamma})^{-1}\mathbb{B}_n^{\Lambda})^{-1}(\mathbb{C}_n^{\beta} - \mathbb{C}_n^{\gamma}(\mathbb{B}_n^{\gamma})^{-1}\mathbb{B}_n^{\beta})\big\}^{-1}, \\
\Sigma_{\gamma,n} &= \big\{\mathbb{B}_n^{\gamma} - \mathbb{B}_n^{\beta}(\mathbb{A}_n^{\beta})^{-1}\mathbb{A}_n^{\gamma} - (\mathbb{B}_n^{\Lambda} - \mathbb{B}_n^{\beta}(\mathbb{A}_n^{\beta})^{-1}\mathbb{A}_n^{\Lambda}) \\
&\qquad \times(\mathbb{C}_n^{\Lambda} - \mathbb{C}_n^{\beta}(\mathbb{A}_n^{\beta})^{-1}\mathbb{A}_n^{\Lambda})^{-1}(\mathbb{C}_n^{\gamma} - \mathbb{C}_n^{\beta}(\mathbb{A}_n^{\beta})^{-1}\mathbb{A}_n^{\gamma})\big\}^{-1}, \\
\Xi_n &= \big\{\mathbb{C}_n^{\Lambda} - \mathbb{C}_n^{\beta}(\mathbb{A}_n^{\beta})^{-1}\mathbb{A}_n^{\Lambda} - (\mathbb{C}_n^{\gamma} - \mathbb{C}_n^{\beta}(\mathbb{A}_n^{\beta})^{-1}\mathbb{A}_n^{\gamma}) \\
&\qquad \times(\mathbb{B}_n^{\gamma} - \mathbb{B}_n^{\beta}(\mathbb{A}_n^{\beta})^{-1}\mathbb{A}_n^{\gamma})^{-1}(\mathbb{B}_n^{\Lambda} - \mathbb{B}_n^{\beta}(\mathbb{A}_n^{\beta})^{-1}\mathbb{A}_n^{\Lambda})\big\}^{-1}.
\end{aligned}$$

Also, for any $t \in (0,\tau)$ define the k-dimensional vectors

$$\Phi_{t,n} = \Big(\widehat{\Delta\Lambda_n}(t_1)1\{t_1 \le t\} \ \ldots \ \widehat{\Delta\Lambda_n}(t_k)1\{t_k \le t\}\Big)'$$

and

$$U_{j,t,n} = (1\{t_1 \le t\} \ \ldots \ 1\{t_k \le t\})',$$

Then the following holds:

Theorem 6.8.1 *Under conditions **C1**-**C7**, the variance estimators $\Sigma_{\beta,n}$, $\Sigma_{\gamma,n}$, and $v_n^2(t) = \Phi_{t,n}'\Xi_n U_{t,n}$ converge in probability to Σ_β, Σ_γ, and $v^2(t)$ $(t \in (0,\tau))$ respectively as $n \to \infty$.*

Proof of Theorem 6.8.1: The proof of Theorem 6.8.1 is based on the arguments given in Parner (1998) and Fang *et al.* (2005).

First, we estimate σ by an empirical version $\sigma_n = (\sigma_{\beta,n}, \sigma_{\gamma,n}, \sigma_{\Lambda_n})$ obtained by replacing θ_0 and P_{θ_0} by $\widehat{\theta}_n$ and $\mathbb{P}_n$ respectively in σ_β, σ_γ, and σ_{Λ_j}. Similar to the proof of Theorem (5.8.1), we can show that σ_n converges in probability to σ uniformly over $\mathcal{H}$, and that its inverse $\sigma_n^{-1} = (\sigma_{\beta,n}^{-1}, \sigma_{\gamma,n}^{-1}, \sigma_{\Lambda_n}^{-1})$ is such that $\sigma_n^{-1}(\mathbf{h})$ converges to $\sigma^{-1}(\mathbf{h})$ in probability.

Recall from 6.31 that, for any $\mathbf{h}_\beta \in \mathbb{R}^q$, the asymptotic variance of $\sqrt{n}\mathbf{h}_\beta'(\widehat{\beta}_n - \beta_0)$ is $[\sigma_\beta^{-1}(\mathbf{h}_\beta, \mathbf{0}_p, 0)]'\mathbf{h}_\beta$. By the consistency of $\hat{\theta}_n$, the dominated convergence theorem and theorem 2.10.6 of van der Vaart and Wellner (1996), it can be shown that $\hat{\sigma}_n(\mathbf{h}_\beta, \mathbf{0}_p, 0)$ is a consisent estimate of $\sigma(\mathbf{h}_\beta, \mathbf{0}_p, 0)$. Hence $[\hat{\sigma}_{\beta,n}^{-1}(\mathbf{h}_\beta, \mathbf{0}_p, 0)]'\mathbf{h}_\beta$ gives a consistent estimate of the asymptotic variance of $\sqrt{n}\mathbf{h}_\beta'(\widehat{\beta}_n - \beta_0)$.

Denote by $\hat{\mathbf{h}}_n = (\hat{\mathbf{h}}_{\beta,n}, \hat{\mathbf{h}}_{\gamma,n}, \hat{h}_{\Lambda,n}) = \sigma_n^{-1}(\mathbf{h}_\beta, \mathbf{0}_p, 0)$. Then $\sigma_n(\hat{\mathbf{h}}_n) = (\mathbf{h}_\beta, \mathbf{0}_p, 0)$, which we can write as

$$\begin{cases}
\sigma_{\beta,n}(\hat{\mathbf{h}}_n) = \mathbf{h}_\beta \\
\sigma_{\gamma,n}(\hat{\mathbf{h}}_n) = \mathbf{0}_p \\
\sigma_{\Lambda_n}(\hat{\mathbf{h}}_n)(u) = 0, \quad \text{for all} \quad u \in [0,\tau].
\end{cases}$$

In particular, let $s = t_1, \ldots, t_k$, in the above system. This yields a system of $(q+p+k)$ equations:

$$\mathbb{D}_n \begin{pmatrix} \hat{\mathbf{h}}_{\beta,n} \\ \hat{\mathbf{h}}_{\gamma,n} \\ \breve{\mathbf{h}}_{\Lambda,n} \end{pmatrix} = \begin{pmatrix} \mathbf{h}_\beta \\ \mathbf{0}_p \\ 0_{s_n} \end{pmatrix} \tag{6.32}$$

where $\breve{\mathbf{h}}_{\Lambda,n} = (\hat{h}_{\Lambda_n}(t_1) \ldots \hat{h}_{\Lambda_n}(t_k))'$. It then follows from directly calculations that $\hat{\mathbf{h}}_{\beta,n} = \Sigma_n \mathbf{h}_\beta$, with Σ_n as given above and therefore, $\mathbf{h}'_\beta \Sigma_n \mathbf{h}_\beta$ is a consistent estimator of the asymptotic variance of $\sqrt{n}\mathbf{h}'_\beta(\widehat{\beta}_n - \beta_0)$ for every $\mathbf{h}_\beta$. It follows that Σ_n is a consistent estimator of Σ.

The consistency of Υ_n is proved along the sames lines.

Let $t \in (0, \tau)$. It follows from the dominated convergence theorem and the consistency of σ_n^{-1} that $\upsilon_n^2(t) = \int_0^t \sigma_{\Lambda,n}^{-1}(\mathbf{0}_q, \mathbf{0}_p, 1\{s \leq t\})\, d\widehat{\Lambda}_n(s)$ converges in probability to $\upsilon^2(t)$. Let $\mathbf{h}_n = (\mathbf{h}_{\beta,n}, \mathbf{h}_{\gamma,n}, h_{\Lambda_n}) = \sigma_n^{-1}(\mathbf{0}_q, \mathbf{0}_p, 1\{s \leq t\})$. Then $\sigma_n(\mathbf{h}_n) = (\mathbf{0}_q, \mathbf{0}_p, 1\{s \leq t\})$, which we can write as

$$\begin{cases} \sigma_{\beta,n}(\mathbf{h}_n) = \mathbf{0}_q \\ \sigma_{\gamma,n}(\mathbf{h}_n) = \mathbf{0}_p \\ \sigma_{\Lambda_n}(\mathbf{h}_n)(u) = 1\{s \leq t\}, \quad \text{for all} \quad u \in [0, \tau] \end{cases} \tag{6.33}$$

In particular, letting $s = t_1, \ldots, t_k$ in (6.33) yields the system

$$\mathbb{D}_n \begin{pmatrix} \mathbf{h}_{\beta,n} \\ \mathbf{h}_{\gamma,n} \\ \mathbf{h}_{\Lambda,n} \end{pmatrix} = \begin{pmatrix} \mathbf{0}_q \\ \mathbf{0}_p \\ U_{t,n} \end{pmatrix}$$

where $\mathbf{h}_{\Lambda,n} = (h_{\Lambda,n}(t_1) \ldots h_{\Lambda,n}(t_k))'$ and $U_{t,n}$ is as defined above. Solving the above system of equation directly yields $\mathbf{h}_{\Lambda,n} = \Xi_n U_{t,n}$. Therefore

$$\begin{aligned} \upsilon_n^2(t) &= \int_0^t \sigma_{\Lambda,n}^{-1}(\mathbf{0}_q, \mathbf{0}_p, 1\{s \leq t\})\, d\widehat{\Lambda}_n(s) \\ &= \sum_{l=1}^{k} \sigma_{\Lambda,n}^{-1}(\mathbf{0}_q, \mathbf{0}_p, 1\{t_l \leq t\}) \widehat{\Delta\Lambda_n}(t_l) 1\{t_l \leq t\} \\ &= \Phi'_{t,n} \mathbf{h}_{\Lambda,n} \end{aligned}$$

and therefore, $\Phi'_{t,n}\Xi_n U_{t,n}$ is a consistent estimator for $\upsilon^2(t)$.

□

Chapter 7

Conclusions

Semiparametric regression models with applications to right censored survival data have gained popularity and an abundant literature has been developed for this class of models. These formulations present theoretical challenges for the presence of infinite dimensional parameters. In this work we have studied two semiparametric models for the survival analysis: a semiparametric mixture model for competing risks and a semiparametric cure model based on transformation models. For the first model, we have adopted the structure proposed by Escarela and Bowater (2008), while for the second model we have taken a specification that generalizes the models established, this specification uses the transformation models and helped to show that the estimators are semiparametric efficient.

The main contribution of this work was to develop a general theory for the NPMLEs in the two models. This theory can easily be used to derive asymptotic results for other semiparametric models in survival analysis. For the two classes of models discussed in this thesis we have identified a set of regularity conditions under which the estimators are consistent, asymptotically normal and efficient. For each model we have presented a representation of the estimator NPML of the cumulative hazard (equations 5.7 and 6.9, respectively) which facilitated and simplified the mathematical derivations of the results presented later. We have extended the techniques developed by Murphy (1994) along the use of the Donsker classes and Glivenko-Cantellil property to demonstrate the consistency of the NPMLEs (Theorem 5.6.1 and Theorem 6.6.1, respectively). The asymptotic normality was demonstrated applying the methodology developed by Murphy (1995) for frailty models, which recommended adopting an analytic function that allows to work with one-dimensional models that pass through the estimator of the model, allowing to see to the space of parameters as a space of functions; this new analytic function allows for an easy way to define operator model information.

From these results it was possible to obtain asymptotic normality for each semipara-

metric model (Theorem 5.7.1, Theorem 5.7.2 and Theorem 6.7.1) and applying the theory established in Bickel et al. (1993) and Tsiatis (2006) we have proved that the efficiency function belongs to the tangent space generated by the score function which shows that the regression parameter estimates are efficient. Finally, the variances were obtained by inverting a linear operator on a space function. The consistent variances for finite and infinite dimensional parameters for each model were obtained following the ideas developed by Parner (1998), Dupuy and Mesbah (2004) and Fang et al. (2005) (Theorem 5.8.1 and Theorem 6.8.1). As the inferences presented in this work have desirable characteristics, which are equivalent to the original Cox model, biostatisticians and statisticians will find in these formulations a convenient, concise and precise form to obtain inferences in models for competing risks and cure models.

The analysis of survival continues presenting gaps. Some issues related to the semiparametric mixture model for competing risks and to semiparametric transformation cure model remain open. Some of these issues can be addressed as follows:

- In the case of the semiparametric mixture model for competing risks, note that the covariate $\mathbf{Z}$ in model 5.3 was assumed to be time independent, as it was in Ng and McLachlan (2003) and Escarela and Bowater (2008). This assumption can be relaxed to accommodate time varying covariates, provided that appropriate regularity conditions are established.

- It would be desirable to extend the conditional failure time model 5.3 to a more flexible class of models, such as the linear transformation models (see Slud and Vonta 2004, for example).

- In the case of the semiparametric transformation cure model is important to develop an estimation process using the EM algorithm for calculate NPMLEs. This result can be achieved using the ideas developed by Peng and Dear (2000) and Sy and Taylor (2000) for the case of the semiparametric proportional hazards cure model. This will allow simulation studies to demonstrate the efficiency of the model in practical situations.

- It may also be interesting to accommodate more complex study designs in the statistical inference for semiparametric mixture models in competing risks data and the semiparametric transformation cure mode; such designs include interval censoring and clustered failure time data.

- The goodness-of-fit tests in these two classes of models should constitute an important direction for future work.

Conclusions en Français

Les modèles de régression semi-paramétriques de durées de vie sont très populaires et une littérature abondante s'est récemment développée pour en étendre le champ d'application. Ces nouvelles formulations présentent des défis théoriques par la présence de paramètres infini-dimensionnels. Dans ce travail nous avons étudié deux modèles semi-paramétriques de l'analyse de survie: un modèle de mélange semi-paramétrique pour les risques concurrents et un modèle semi-paramétrique avec fraction inmune basé sur les modèles de transformation. Pour le premier modèle nous avons adopté la formulation proposée par Escarela et Bowater (2008) et pour le deuxième modèle une formulation très générale a été adoptée.

La principale contribution de ce travail est de développer une théorie générale pour les estimateurs dits du maximum de vraisemblance non-paramétrique (ou NPMLE) dans ces deux modèles. La théorie présentée peut être facilement utilisée pour obtenir des résultats asymptotiques pour d'autres modèles semi-paramétriques de l'analyse de survie. Pour les deux classes de modèles discutées dans ce travail, nous avons identifié un ensemble de conditions de régularité sous lesquelles les estimateurs sont consistants, asymptotiquement gaussiens et efficaces. Pour chaque modèle, une représentation intégrale de l'estimateur NPML du paramètre fonctionnel a été obtenue. Cette représentation facilite et simplifie les calculs mathématiques des résultats postérieurs. Nous avons utilisé et adapté les techniques développées par Murphy (1994, 1995) pour le modèle de fragilité. Nous nous sommes appuyés sur des outils de la théorie des processus empiriques pour obtenir nos résultats. En particulier, nous avons montré, outre l'existence des estimateurs NPML, leur consistance, leur normalité asymptotique, et leur efficacité au sens semi-paramétrique. Ces derniers résultats font appel à la théorie de l'efficacité dans les modèles semi-paramétriques (voir Bickel et al. (1993) et Tsiatis (2006)). Enfin, nous avons proposé des estimateurs convergents pour les variances asymptotiques de nos estimateurs. L'inférence peut ensuite déboucher sur des outils appliqués (tests d'hypothèses, calculs d'intervalles de confiance par exemple), d'intérêt pour les biostatisticiens par exemple.

De nombreuses questions relatives à la modélisation semi-paramétrique des durées de vie avec risques concurrents et fraction immune restent ouvertes. Nous en énumérons quelques unes:

- Dans le cas du modèle de mélange semi-paramétrique pour des risques concurrents, notons que la covariable $\mathbf{Z}$ dans le modèle (5.4) a été considérée comme une variable indépendante du temps, comme il a été fait par Ng et McLachlan (2003) et Escarela et Bowater (2008). Toutefois, cette supposition peut être assouplie, c'est à dire nous pouvons permettre que le modèle (5.4) inclue des variables dépendant du temps, ce qui peut se faire au prix d'hypothèses de régularité supplémentaires.

  Ce même modèle pourrait être étendu au cas où la distribution des temps d'évènements est donnée par un modèle de transformation linéaire, incluant le modèle de Cox adopté dans ce travail.

- Il serait aussi intéressant d'incorporer dans nos analyses des situations de censure plus complexes: censures par intervalle par exemple, ou des design plus compliqués: présence de clusters familiaux par exemple.

- Dans le cas du modèle semi-paramétrique de transformation avec fraction immune, une piste importante de travail futur est fournie par l'étude des aspects algorithmiques de l'estimation: mise en oeuvre d'un algorithme d'estimation, étude, par simulation, des propriétés des estimateurs proposés pour des tailles d'échantillons finies. Il serait souhaitable faire une extension du modèle qui permet incorporer censures par intervalle.

- Il serait également intéressant de considérer les problèmes d'ajustement pour ces modèles de plus en plus complexes.

Bibliography

[1] P.K. Andersen and R.D. Gill. Cox's regression model for counting processes: A large sample study. *Ann. Statist.*, 10(4):1100–1120, 1982.

[2] V. Bagdonavicius and M.S. Nikulin. Generalized proportional hazards model based on modified partial likelihood. *Lifetime Data Analysis*, 5(4):329–350, 1999.

[3] V. Bagdonavicius and M.S. Nikulin. *Accelerated Life Models : Modeling and Statistical Analysis*. Chapman & Hall, Boca Ratón (Florida), 2002.

[4] R. Bender, T. Augustin, and M. Blettner. Generating survival times to simulate cox proportional hazards models. *Statist. in Medicine*, 24:1713–1723, 2005.

[5] J. Berkson and R.P. Gage. Survival curve for cancer patients following treatment. *J. Amer. Statis. Assoc.*, 47(259):501–515, 1952.

[6] P. Bickel, C. Klaassen, Y. Ritov, and J. Wellner. *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins University Press, Baltimore, 1993.

[7] J.M. Boag. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J. R. Statis. Soc. Ser. B*, 11(1):15–53, 1949.

[8] N. Breslow. Covariance analysis of censored survival data. *Biometrics*, 30(1):89–99, 1974.

[9] A.B. Cantor and J.J. Shuster. Parametric versus non-parametric methods for estimating cure rates based on censored survival data. *Statist. in Medicine*, 11(7):931–937, 1992.

[10] J.F. Carrire. Removing cancer when it is correlated with other causes of death. *Biometrical Journal*, 37(3):339–350, 1995.

[11] I.S. Chang, C.A. Hsuing, M.C. Wang, and C.C. Wen. An asymptotic theory for the nonparametric maximum likelihood estimator in the cox gene model. *Bernoulli*, 11(5):863–892, 2005.

[12] K. Chen, Z. Jin, and Z. Ying. Semiparametric analysis of transformation models with censored data. *Biometrika*, 89(3):659–668, 2002.

[13] S.C. Cheng, F.P. Fine, and L.J. Wei. Prediction of cumulative incidence function under the proportional hazards model. *Biometrika*, 82(4):835–845, 1995.

[14] S.C. Cheng, L.J. Wei, and Z. Ying. Analysis of transformation models with censored data. *Biometrics*, 54(1):219–228, 1998.

[15] K.C. Choi and X. Zhou. Large sample properties of mixture models with covariates for competing risks. *J. Multivar. Anal*, 82(2):331–366, 2002.

[16] D. Clayton and J. Cuzick. Multivariate generalizations of the proportional hazards model (with discussion). *J. R. Statis. Soc.. Ser. A*, 148(2):82–117, 1985.

[17] D. Collet. *Modelling survival data in medical research*. Chapman & Hall, London, 1994.

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

[19] D. R. Cox and D. Oakes. *Analysis of Survival Data*. Chapman & Hall, London, 1984.

[20] D. R. Cox and E.J. Snell. *Analysis of Binary Data*. Chapman & Hall, London, 1989.

[21] J. Cuzick. Rank regression. *Ann. Statist.*, 16(4):1369–1389, 1988.

[22] A.P. Dempster, N.M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *J. R. Stat. Soc. Ser. B*, 39(1):1–38, 1977.

[23] J.-F. Dupuy and G. Escarela. Modélisation de risques concurrents par un modèle de mlange semi-paramétrique. *Comptes Rendus Mathematique*, 34(10):641–644., 2007.

[24] J.-F. Dupuy, I. Grama, and M. Mesbah. Asymptotic theory for the Cox model with missing time-dependent covariate. *Ann. Statist.*, 34(2):903–924, 2006.

[25] J.-F. Dupuy and M. Mesbah. Estimation of the asymptotic variance of semiparametric maximum likelihood estimators in the Cox model with a missing time-dependent covariate. *Comm. Statist. Theory Methods*, 33(6):1385–1401, 2004.

[26] B. Efron. The efficiency of cox's likelihood function for censored data. *J. Amer. Stat. Assoc.*, 72(359):557–565, 1977.

[27] R.C. Eldant-Johnson and N.L. Johnson. *Survival models and data analysis*. Wiley, New York, 1980.

[28] G. Escarela and R. Bowater. Fitting a semi-parametric mixture model for competing risks in survival data. *Comm. Statist. Theory Methods*, 37(2), 2008.

[29] H.-B. Fang, G. Li, and J. Sun. Maximum likelihood estimation in a semiparametric logistic/proportional-hazards mixture model. *Scand. J. Statist.*, 32(1):59–75, 2005.

[30] V.T. Farewell. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, 38(4):1041–1046, 1982.

[31] V.T. Farewell. Mixture models in survival analysis: are they worth the risk? *Can. J. Statist.*, 14(3):257–262, 1986.

[32] J.P Fine. Analysing competing risks data with transformation models. *J. R. Stat. Soc. Ser. B*, 61(4):817–830, 1999.

[33] T. R. Fleming and D. P. Harrington. *Counting processes and survival analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, 1991.

[34] J.J. Gaynor, E.J. Feuer, C.C Tan, D.H. Wu, C.R. Little, D.J. Straus, B.D. Clarkson, and M.F. Brennan. On the use of cause-specific failure and conditional failure probabilities: Examples from clinical oncology data. *J. Amer. Statist. Assoc.*, 88(2):218–241, 1993.

[35] M.E. Ghitany, R. A. Maller, and S. Zhou. Exponential mixture models with long-term survivors and covariates. *J. Multivar. Anal*, 49(422):400–409, 1994.

[36] B. Grun and F. Leisch. Identifiability of finite mixtures of multinomial logit models with varying and fixed effects. *Technical Report, Department of Statistics,University of Munich*, (024):329–350, 2008.

[37] A. Hernández Quintero, J.-F. Dupuy, and G. Escarela. Analysis of a semiparametric mixture model for competing risks. *Ann. Inst. Statis. Math.*, 2009.

[38] J. D. Holt. Competing risk analyses with special reference to matched pair experiments. *Biometrika*, 65(1):159–165, 1978.

[39] C. Huber and J.P. Lecoutre. Estimation fonctionnelle dans les modèles de durèe. *Analyse statistique des durèes de vie*, pages 59–120, 1989.

[40] S. Johansen. An extension of cox's regression model. *Int. Stat. Rev.*, 51(2):165–174, 1983.

[41] N.L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions.Volume 1.* Wiley Series in Probability and Statistics, New York, 1994.

[42] D.R. Jones, R.L. Powles, D. Machin, and Sylvester R.J. On estimating the proportion of cured patients in clinical studies. *Biometrie-Praximetrie*, 21:1–11, 1981.

[43] J. D. Kalbfleisch and R. L. Prentice. Marginal likelihoods based on cox's regression and life model. *Biometrika*, 60(2):267–278, 1973.

[44] J.D. Kalbfleisch and R.L. Prentice. *The Statistical Analysis of Failure Time Data*. Willey, New York, 1980.

[45] J. P. Klein and R. Bajorunaite. Inference for competing risks. In *Advances in survival analysis*, volume 23 of *Handbook of Statist.*, pages 291–311. Elsevier, Amsterdam, 2004.

[46] J.P. Klein. Semiparametric estimation of random effects using the cox model based on the em algorithm. *Biometrics*, 48(3):795–806, 1992.

[47] J.P. Klein and M.L. Moeschberger. *Survival Analysis. Techniques for Censored and Truncated Data*. Statistics for Biology and Healt. Springer, New York, 1997.

[48] S. Konishi and G. Kitagawa. *Information Criteria and Statistical Modeling*. Springer, New York, 2008.

[49] M. R. Kosorok and R. Song. Inference under right censoring for transformation models with a change-point based on a covariate threshold. *Ann. Statist.*, 35(3):957–989, 2007.

[50] A.Y.C. Kuk. A semiparametric mixture model for the analysis of competing risks data. *Australian & New Zealand Journal of Statistics*, 34(2):169–180, 1992.

[51] A.Y.C. Kuk and C.H. Chen. A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, 79(3):531–541, 1992.

[52] S. Kullback and R. A. . Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 1951.

[53] M.G. Lagakos, C.J Sommer, and M. Zelen. Semi-markov models for partially censored data. *Biometrika*, 65(2):311–317, 1978.

[54] M. G. Larson and G. E. Dinse. A mixture model for the regression analysis of competing risks data. *J. Roy. Statist. Soc. Ser. C*, 34(3):201–211, 1985.

[55] J.F. Lawlees. *Statistical Models and Methods for Lifetime Data*. Wiley Series in Probability and Mathematical Statistics. Wiley, New York, 1982.

[56] Y. Lo, J.M. Taylor, W.H. McBride, and Withers H.R. The effect of fractionated doses of radiation on mouse spinal cord. *Int. J. Radiat. Oncol. Biol. Phys.*, 27(2):309–317, 1993.

[57] W. Lu. Maximum likelihood estimation in the proportional hazards cure model. *Ann. Inst. Statist. Math.*, 60(3):545–574, 2008.

[58] W. Lu and Z. Ying. On semiparametric transformation cure models. *Biometrika*, 91(2):331–343, 2004.

[59] S. Ma and Kosorok M.R. Penalized log-likelihood estimation for partly linear transformation models with current status data. *Ann. Statist.*, 33(5):2256–2290, 2005.

[60] R. A. Maller and S. Zhou. Estimating the proportion of immunes in a censored sample. *Biometrika*, 79(4):731–739, 1992.

[61] R.A Maller and S. Zhou. *Survival Analysis with Long-Term Survivors*. Wiley, New York, 1996.

[62] R.A. Maller and X. Zhou. Analysis of parametric models for competing risks. *Statistica Sinica*, 12:725–750, 2002.

[63] T. Martinussen and T. H. Scheike. *Dynamic regression models for survival data*. Statistics for Biology and Health. Springer, New York, 2006.

[64] W.Q. Meeker and L.A. Escobar. *Statistical Methods for Reliability Data*. Wiley Series in Probability and Statistics, New York, 1998.

[65] S. A. Murphy. Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.*, 22(2):712–731, 1994.

[66] S. A. Murphy. Asymptotic theory for the frailty model. *Ann. Statist.*, 23(1):182–198, 1995.

[67] S. A. Murphy, A. J. Rossini, and A. W. Van der Vaart. Maximum likelihood estimation in the proportional odds model. *J. Amer. Statist. Assoc.*, 92(439):968–976, 1997.

[68] M. Naskar, K. Das, and J.G. Ibrahim. A semiparametric mixture model for analyzing clustered competing risks data. *Biometrics*, 61(3):729–737, 2005.

[69] S. K. Ng and G. J. McLachlan. An em-based semi-parametric mixture model approach to the regression analysis of competing-risks data. *Stat. Med.*, 22:1097–1111, 2003.

[70] G.G. Nielsen, R.D. Gill, P.K. Andersen, and T.I.A. Srensen. A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist*, 19(1):25–43, 1992.

[71] S.F. Pack and Morgan B.J. T. A mixture model for interval-censored time-to-response quantal assay data. *Biometrics*, 46(3):749–757, 1990.

[72] E. Parner. Asymptotic theory for the correlated gamma-frailty model. *Ann. Statist.*, 26(1):183–214, 1998.

[73] Y. Peng. Fitting semiparametric cure models. *Comput. Statist. Data Anal.*, 41(3-4):481–490, 2003.

[74] Y. Peng and K.B.G. Dear. A nonparametric mixture model for cure rate estimation. *Biometrics*, 56(1):237–243, 2000.

[75] Y. Peng, K.B.G Dear, and J.W Denham. A generalized f mixture model for cure rate estimation. *Statist. in Med.*, 17(8):813–830, 1998.

[76] R.L. Prentice, J.D. Kalbfleisch, A.V. Peterson, N. Flournoy, V.T. Farewell, and Breslow N.E. The analysis of failure times in the presence of competing risks. *Biometrics.*, 34(4):41–54, 1978.

[77] E. V. Slud and F. Vonta. Consistency of the NPML estimator in the right-censored transformation model. *Scand. J. Statist.*, 31(1):21–41, 2004.

[78] L.B Spivey and A.J Gross. Concomitant information in competing risk analysis. *Biometrical Journal*, 33(4):419 – 427, 2007.

[79] R. Sposto, H.N Sather, and S.A. Baker. A comparison of tests of the difference in the proportion of patients who are cured. *Biometrics*, 48(1):87–99, 1992.

[80] J. P. Sy and J. M. G. Taylor. Estimation in a Cox proportional hazards cure model. *Biometrics*, 56(1):227–236, 2000.

[81] J. M. G. Taylor. Semi-parametric estimation in failure time mixture models. *Biometrics*, 51(3):899–907, 1995.

[82] A. Tsiatis. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences of the United States of America*, 72(1):20–22, 1975.

[83] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1998.

[84] A. W. van der Vaart and J. A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996.

[85] V.G. Voinov and M.S. Nikulin. *Unbiased estimators ann their applications. Volume 1: Univariate Case*. Mathematics and Its Applications, Kluwer Academic Publisher, Dordrecht, 1993.

[86] G.A. Whitmore. First-passage-time models for duration data: Regression structures and competing risks. *J. R. Stat. Soc. Ser. D*, 35(2):207–219, 1986.

[87] K. Yamaguchi. Accelerated failure-time regression models with a regression model of surviving fraction: an application to the analysis of "permanent employment" in japan. *J. Amer. Statist. Assoc.*, 87(418):284–292, 1992.

[88] D. Zeng and D. Y. Lin. Maximum likelihood estimation in semiparametric regression models with censored data. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 69(4):507–564, 2007.

[89] D. Zeng and D. Y. Lin. Semiparametric transformation models with random effects for recurrent events. *J. Amer. Statist. Assoc.*, 102(477):167–180, 2007.

[90] D. Zeng, D. Y. Lin, and G. Yin. Maximum likelihood estimation for the proportional odds model with random effects. *J. Amer. Statist. Assoc.*, 100(470):470–483, 2005.

[91] D. Zeng and D.Y. Lin. Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*, 93(3):627–640, 2006.

[92] D. Zeng and D.Y. Lin. A general asymptotic theory for maximum likelihood estimation in semiparametric regression models with censored data. *Technical Report. University of North Carolina, Chapell Hill*, 2007.

[93] D. Zeng, D.Y. Lin, and X. Lin. Semiparametric transformation models with random effects for clustered failure time data. *Statistica Sinica*, 18(1):355–377, 2008.