FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Jaromír Macoun

Competing risk models in survival analysis

Department of Probability and Mathematical Statistics

Supervisor of the master thesis:	doc. RNDr. Arnošt Komárek, Ph.D.
Study programme:	Probability, Mathematical Statistics and Econometrics

Prague 2024

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In .... date ....

Author's signature

I would like to express my gratitude to my supervisor doc. Komárek for his advice and comments without which I would have had a problem to finish the thesis. I would also like to thank my family and my girlfriend for their endless support not only during the writing of the thesis but throughout my studies. Finally, I would like to thank Maja Pohar Perme and Erik Langerholc who inspired me during my stay abroad in certain passages of the thesis.

Title: Competing risk models
in survival analysis

Author: Jaromír Macoun

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Arnošt Komárek, Ph.D., Department of Probability and Mathematical Statistics

Abstract: We study the extension of methods from classical survival analysis to competing risks. These methods can be used to analyse time-to-event data. Firstly, we establish notation, define fundamental concepts, and present basic theorems and properties. The second chapter explores semi-parametric methods for estimating the cumulative incidence function. We compare two methods of estimation: the first treats competing events as censored, while the second takes competing events into account. At the end of the chapter, we prove the asymptotic distribution of the estimator of the cumulative incidence function. Next, we present semi-parametric regression methods for estimating cause-specific and subdistribution hazards. Generalisations of the Cox model are used to estimate regression parameters. We introduce proofs of the martingale property for the subdistribution hazard case with complete data. Lastly, we propose a small simulation study to assess the efficiency of the presented nonparametric estimates. Different scenarios with constant cause-specific hazards are simulated and visualised. Additionally, there is one more simulation study for semiparametric estimation methods. Two different Cox models with two covariates for cause-specific hazard are assumed.

Keywords: competing risks, cumulative incidence function, cause-specific hazard, subdistribution hazard, Cox model

Contents

Notation

T	Time to event
D	Indicator of the type of event
C	Time to censoring
X	Time to event, or censoring
$\overline{N}_j(t)$	Counting process for the j-th event
$\overline{Y}(t)$	At risk process
$\mathbf{1}\{.\}$	Indicator function
$o_p(1)$	Term negligible in probability
$S(t)$	Survival function
$\Lambda_j^{CS}(t)$	Cumulative cause-specific hazard function of the j-th event
$\lambda_j^{CS}(t)$	Cause-specific hazard function of the j-th event
$\lambda_j^{SD}(t)$	Subdistributional function of the j-th event
$CIF_j(t)$	Cumulative incidence function of the j-th event
$\widehat{KM}(t)$	Kaplan-Meier estimator of the survival function
$\xrightarrow{D[0,\tau]}$	Weak convergence in the $D[0,\tau]$ space

Introduction

In survival analysis, we look at either time to event or time to censoring. This type of analysis is commonly used in medical data, for example time to death. However, sometimes there are several types of death and in clinical practice it is necessary to distinguish between them. We have an event of interest and a competing event. The basic assumption is that if one type of event occurs, then the other cannot be observed. This assumption contradicts censoring, because when censoring is being observed , in fact any other event can happen. Therefore, it is not correct approach treating competing events as censoring. Multi-state models could be used for this situation, because the competing risk problem is just a special case of the multi-state model. In the competing risk setting model has an initial state and k absorbing states. As a consequence, results from multi-state models can be used. However,the problem can be approached from the opposite side. We will present and derive all definitions and properties as a generalisation of classical survival analysis rather than as a special case of multi-state models.

In the first chapter we will summarise the basic knowledge of classical survival analysis and define the most important quantities for competing risks. Explanations and relationships of the definitions are given to familiarise the reader with the problem.

The next chapter focuses on non-parametric methods for estimating the cumulative incidence function. The first part of the chapter discusses two possible estimators, one that takes competing events into account and the second that treats competing events as censoring. We show that the second estimator is biased and that there is an inequality between these estimators. The main result of this chapter is the derivation of the limiting process of the first estimator. To prove this theorem, the estimator is rewritten as a martingale and then the central limit theorem for martingales is used.

Semi-parametric methods are discussed in the third chapter. First, a Cox regression model for cause-specific hazards is presented. However, we focus more on modelling sub-distribution hazards. Three types of data are considered and a generalisation of the Cox model is presented for each type. In addition, the martingale property is derived for two out of three situations. This property is crucial for deriving asymptotic properties.

In the last chapter we present two simulations, one for nonparametric methods and one for semiparametric methods. There are two different ways of generating simulated data. In the non-parametric simulation we study the efficiency of estimators of a cumulative incidence function. Similarly, we simulate data given by the Cox cause-specific hazard model. We have the regression model with a numerical covariate and a binary variable. We compare the estimators according to different scenarios such as different percentage of censoring, sample size or different setting of initial values of the cause-specific hazard. All simulations were calculated using R Core Team [2023].

1. Basic Concepts and Definitions

This chapter provides motivation and a brief summary of classical survival analysis. Basic definitions and theorems are crucial for understanding competing risk models as an extension of survival analysis. Secondly, motivation for competing risk problems is introduced. Main definitions such as a cumulative incidence function, cause-specific, and subdistribution hazard are presented with stated notation. At the end of the chapter, elementary properties and relations among defined quantities are derived, which helps to understand the meaning of the definitions.

1.1 Introduction to Survival Analysis

Survival analysis deals with subjects for which the time until a specific event is observed. We often refer to the time until an event as the time to failure. In practice, such data formats can be found in various fields such as medicine, industry, insurance, and many other disciplines. For instance, we can track patients with transplanted hearts and observe the time of patient's death. Similar example can be found in industry where we monitor a machine and the time until it fails. It's not always necessary to track time units; we can also monitor the energy consumed by a machine or the number of products produced before it fails. The issue would lead to a non-negative discrete random variable in the latter case. Theory has also been developed for this case; however, we will focus only on continuous variables in this work. Therefore, we investigate a non-negative continuous random variable T, which we will call the "time to failure."

From the continuity and non-negativity of a random variable, it follows that $P(T = 0) = 0$. For example, the failure time distribution T can be determined by its cumulative distribution function. This function specifies the probability of a failure occurring within the time interval $t \in [0, \infty)$. For negative t, the cumulative distribution function is trivially zero. Due to the continuity and non-negativity of the random variable, we will consider $t \in (0, \infty)$. From an interpretative perspective, we are more interested in the probability that a patient survives beyond time t. That is the reason why we work with the survival function rather than the cumulative distribution function.

Definition 1 (Survival function)**.** *The* survival function *of the random variable T with the cumulative distribution function F is defined as $S(t) = 1 - F(t) = P(T > t)$.*

The distribution of T can be determined by the cumulative distribution function, density, or survival function. A non-negative continuous random variable can also be characterised by a hazard function or a cumulative hazard function. The relevant definitions and a theorem summarising the relationships of these functions with the survival function are provided below.

Definition 2 (Hazard function)**.** *Let T be a continuous non-negative random variable. The* hazard function $\lambda(t)$ *is defined as:*

$$\lambda(t) = \lim_{h \downarrow 0} \frac{P(t \leq T < t + h | T \geq t)}{h}.$$

Definition 3 (Cumulative hazard function)**.** *Let T be a continuous non-negative random variable. The* cumulative hazard function $\Lambda(t)$ *is defined as:*

$$\Lambda(t) = \int_0^t \lambda(s) ds.$$

Theorem 1. *Let T be a non-negative continuous random variable with a cumulative distribution function F and a survival function $S = 1 - F$, or a density function f. Then the following relationships hold:*

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

ii $\Lambda(t) = \int_0^t \frac{dF(s)}{S(s)}$

iii $S(t) = \exp\left(-\Lambda(t)\right)$

Proof. Proofs are provided in Kalbfleisch and Prentice [2011, part 1.2.1]. □

Therefore, we observe n subjects for a certain period and record the failure times. In practice, we may not always wait for the failure time for every subject, or a patient may stop being monitored for other reasons. So, we observe the failure time for some individuals and the time of leaving the study for others (censoring). We do not have a random sample of $T_1, \ldots, T_n$ from a distribution determined by a random variable T; instead, we observe two-dimensional random vectors $(X_1, \delta_1)^T, \ldots, (X_n, \delta_n)^T$.

The random variable X_i represents the time either until the failure or censoring of subject i, and the binary random variable δ_i serves as an indicator of whether the failure time was observed ($\delta_i = 1$) or not ($\delta_i = 0$).

The random variable X_i is defined as $\min(T_i, C_i)$, where T_i is the event time and C_i is the censoring time of the i-th individual. The random variable δ_i is defined as $\mathbf{1}(T_i \leq C_i)$, indicating whether the event occurred before censoring.

For further work, it is necessary to introduce a certain assumption of independence between the event time T and the censoring time C. These random variables without an index are considered generic random variables from which observations are generated.

Definition 4 (Independence Censoring Condition)**.** *Let T and C be non-negative random variables, where the random variable X is the minimum of T and C. If the following condition holds, C satisfies the independence censoring condition.*

$$\lambda(t) = \lim_{h \downarrow 0} \frac{P(t \leq T < t + h | X \geq t)}{h}$$

Under the condition of independent censoring, one can then construct estimates of the characteristics of the random variable T that have already been introduced above. Before presenting these estimates, we will introduce notation in the context of counting processes and martingales since this technique is important for later proving properties. The counting process $N_i(t) = \mathbf{1}\{T_i \leq t; \delta = 1\}$ is a right-continuous process that takes on the value 0, and at the time of failure, it jumps to the value 1. In other words, this process provides information about whether an event occurred by time t for the i-th subject. Another process that is used is $Y_i(t) = \mathbf{1}\{X_i \geq t\}$, called at risk process. This process takes on the value 1 if neither an event nor censoring has occurred for the subject. These two processes form the generator of the sigma-algebra system for which we will want to construct a martingale. The form of this filtration for all observations is $\mathcal{F}_t = \sigma\{N_i(u), Y_i(u+), 0 \leq u \leq t, i = 1, \ldots, n\}$. The counting process is a non-negative right-continuous submartingale. Therefore, the existence of a compensator for the counting process follows from the Doob-Meyer decomposition. This procedure is summarised in the following theorem.

Theorem 2. *Let $A(t) = \int_0^t Y(u) d\Lambda(u)$. Then $A(t)$ is a right-continuous predictable process. It holds that the process $M(t) = N(t) - A(t)$ is a martingale with respect to the filtration $\mathcal{F}_t$, if and only if the condition of independent censoring in Definition 4 is satisfied.*

Proof. For detailed proof see Fleming and Harrington [2011, Theorem 1.3.1]. □

Next, let's assume that we are observing a two-dimensional random sample $(X_1, \delta_1), \ldots, (X_n, \delta_n)$, as defined earlier, and the condition of independent censoring is satisfied. We denote $\overline{N}(t) = \sum_{i=1}^n N_i(t)$, which represents the sum of subjects for whom failure (event) occurred by time t. Similarly, $\overline{Y}(t) = \sum_{i=1}^n Y_i(t)$ indicates the number of individuals for whom neither an event nor censoring occurred by time t. Below, some estimation methods are summarised.

Definition 5 (Kaplan-Meier Estimator of Survival Function)**.** *The Kaplan-Meier estimator of the survival function is defined:*

$$\widehat{S}(t) = \prod_{t_j \leq t} (1 - \hat{\lambda}_j) = 1 - \sum_{t_j \leq t} \widehat{S}(t_j-) \frac{\Delta \overline{N}(t_j)}{\overline{Y}(t_j)},$$

where t_j are the distinct event times obtained from the data for n individuals, and $\hat{\lambda}_j = \frac{\Delta \overline{N}(t_j)}{\overline{Y}(t_j)}$ represents the estimated intensity at time t_j.

Definition 6 (Nelson-Aalen Estimator of Cumulative Hazard)**.** *The Nelson-Aalen estimator of the cumulative hazard is defined as follows:*

$$\hat{\Lambda}(t) = \int_0^t \frac{d\overline{N}(u)}{\overline{Y}(u)}$$

Assume that we have maximum time $\tau > 0$. For more details and proofs see Andersen et al. [2012, part IV.1.2]. It is well known that Nelson-Aalen and Kaplan-Meier estimators are uniformly consistent if we assume the condition of independent censoring (Definition 4) and with some assumption on the at-risk process. In other words, expressions

$$\sup_{t \in [0,\tau]} |\hat{\Lambda}(t) - \Lambda(t)| \text{ and } \sup_{t \in [0,\tau]} |\widehat{S}(t) - S(t)|$$

converge to zero in probability.

The Kaplan-Meier estimator can also be viewed from a different perspective. Here, we follow the approach presented in the work by Efron [1967]. We have n subjects, and events or censoring occur sequentially at times $x_1 < x_2 < \cdots < x_n$. Implicitly, this assumes that at each time x_i, either an event or censoring is observed, and two events are never observed simultaneously. Without censoring, the Kaplan-Meier estimator would be a step function, decreasing by $1/n$ at each x_i. However, censoring is prevalent in the data in practice, so the individual step sizes are adjusted based on the number of subjects censored. The procedure is as follows. Assign a probability of $1/n$ to each x_i. When the first censoring occurs at time x_{i_1}, remove the probability assigned to this time (which is now $1/n$) and distribute it equally among the remaining observations $x_{i_1+1}, \ldots, x_n$. This redistribution of probability $1/n$ is performed because the subject did not experience failure and may fail in the future. Thus, the remaining times $x_{i_1+1}, \ldots, x_n$ now have probabilities of

$$\frac{1}{n} + \frac{1}{n(n-i_1)} = \frac{1}{n}\left(1 + \frac{1}{n-i_1}\right).$$

This inductive process helps to understand how the Kaplan-Meier estimator deals with censored observations. Moreover, from just two steps, it is apparent that the time x_i for $i \in \{1, \ldots, n-1\}$ has a probability determined by the expression

$$\frac{1}{n}\prod_{j=1}^{i-1}\left(1 + \frac{1}{n-j}\right)^{1-\Delta\overline{N}(x_j)}, \tag{1.1}$$

if no censoring occurred at time x_i. For time x_n, the formula in (1.1) is valid whether censoring or an event occurred. To estimate $P(T > t)$, you need to sum the probabilities of all times for which $x_i > t$. The following proposition relates this construction to the Kaplan-Meier estimator and offers another perspective on how to interpret this estimator.

Proposition 3. *Let us have times $x_1 < x_2 \ldots x_{n-1} < x_n$, then the following equality holds for $t \in (x_{k-1}, x_k]$*

$$\hat{S}(t-) = \frac{\overline{Y}(t)}{n}\frac{1}{\hat{G}(t-)},$$

where $\hat{G}(t)$ is Kaplan-Meier estimate of censoring event.

Proof. Let's make some algebraic operations with the Kaplan-Meier estimate.

$$\hat{S}(t-) = \prod_{x_i<t}\left(\frac{\overline{Y}(x_i)-1}{\overline{Y}(x_i)}\right)^{\Delta\overline{N}(x_i)} = \prod_{x_i<t}\left(\frac{\overline{Y}(x_i)}{\overline{Y}(x_i)-1}\right)^{1-\Delta\overline{N}(x_i)}\prod_{x_i<t}\left(\frac{\overline{Y}(x_i)-1}{\overline{Y}(x_i)}\right) =$$

$$= \frac{1}{\hat{G}(t-)}\frac{\overline{Y}(x_1)-1}{\overline{Y}(x_1)}\frac{\overline{Y}(x_2)-1}{\overline{Y}(x_2)}\cdots\frac{\overline{Y}(x_k)}{\overline{Y}(x_{k-1})}$$

Last expression is actually telescopic product since $\overline{Y}(x_1)-1 = \overline{Y}(x_2)$, so product is simplified. For the end of the proof, we need to realise that $\overline{Y}(x_k) = \overline{Y}(t)$. □

Corollary. We can see that the weight of observation x_i is in the notation of Proposition 3 indeed

$$\frac{1}{n\hat{G}(t-)}.$$

The different form of a Kaplan-Meier estimate should give readers an idea of how the estimate deals with censored data. This approach and reasoning will be used subsequently in competing risk models.

The expression of the Kaplan-Meier estimate in Proposition 3 can be imagined as a classic empirical survival function (analogy of empirical distribution function). If there is no censoring we could estimate for $x_{k-1} \leq t < x_k$ probability $P(T > t)$ just by $(n-k+1)/n$. If censoring occurs in data, we do not have the same weight for all observations. The size of each weight is expressed by Proposition 3. As a consequence, from the other point of view, we can see

$$\hat{S}(t-) = \sum_{j=k}^{n} \frac{1}{n\hat{G}(t-)} = \frac{\overline{Y}(t)}{n\hat{G}(t-)}.$$

The Cox proportional hazard model is helpful in evaluating the joint effect of more than one covariate. The advantage of this approach is that we are not directly assuming any distribution, but we have one crucial assumption, which must be satisfied, the proportional hazard assumption. However, proportional hazards are more relaxed than specifying a distribution. For completeness of this thesis, we'd like to present just a brief intro into the theory of the Cox model. For more details, see Andersen et al. [2012, part VII.2.1 and VII.2.2].

Assume that the hazard depends on some set covariates. Let us have an explanatory variable vector (stochastic process) $\boldsymbol{Z}(t)$. We need to modify the independent censoring condition from the Definition 4 into

$$\begin{aligned}\lambda(t|\boldsymbol{Z}) &= \lim_{h\downarrow 0} \frac{P(t \leq T < t+h | T \geq t, \boldsymbol{Z}(t))}{h} \\ &= \lim_{h\downarrow 0} \frac{P(t \leq T < t+h | T \geq t, C \geq t, \boldsymbol{Z}(t))}{h}.\end{aligned}$$

Let us have a regression parameter (p-dimensional vector) $\boldsymbol{\beta}$. Assume that there is a parameter value $\boldsymbol{\beta^0}$ and a baseline hazard $\lambda_0(t)$ such that it holds

$$\lambda_{\boldsymbol{Z}}^{CS}(t) = \lambda_0^{CS}(t) \exp\left(\boldsymbol{Z}^T(t)\boldsymbol{\beta^0}\right). \tag{1.2}$$

Assume that $\lambda_0(t)$ is some unknown unspecified hazard function. Furthermore, assume independence across different subjects. Based on this assumption, we can construct the partial likelihood.

Definition 7. *The function*

$$\mathbf{L}(\boldsymbol{\beta}) = \prod_{i=1}^{n} \prod_{s>0} \left(\frac{Y_i(s) \exp\left(\boldsymbol{\beta}^T \boldsymbol{Z_i}(s)\right)}{\sum_{l=1}^{n} Y_l(s) \exp\left(\boldsymbol{\beta}^T \boldsymbol{Z_l}(s)\right)} \right)^{\Delta N_i(s)}$$

is called the partial likelihood, *where* $\Delta N_i(t)$ *denotes the change in the counting process at time* t.

Maximizing this function provides estimate $\hat{\boldsymbol{\beta}}$ for the regression coefficients. We skip the part with log partial likelihood and score function since we will present analogous derivations for cause-specific and subdistribution hazard in sections 3.1 and 3.2. It is good to remember that with some additional assumptions expression $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)$ converges in distribution to $p-$dimensional normal distribution with zero mean and some covariance matrix I^{-1}. With this result confidence intervals and statistical tests could be easily constructed.

1.2 Introduction to competing risks

This section will generalise survival analysis problems to cases where an event can occur for multiple reasons. To illustrate, imagine a doctor monitoring the time of a patient's death after a heart transplant. However, death can occur for various reasons, such as pulmonary embolism, arrhythmia, infection, or many other causes. In addition to monitoring the time to event T, we also consider the type of event. Suppose there are K different event types, where K is a finite natural number. A fundamental assumption is that no other type can occur once one type of event occurs.

If we are interested in only one event, we could think of it as standard survival analysis, as we can treat the other reasons for events as censoring. In this case, we might violate some assumptions. In survival analysis, we assume that a censored subject can still experience an event. For example, a patient who has moved (censoring time) and is no longer being monitored, may experience an event after the censoring time. If we consider competing risks as censoring, it contradicts the assumption that no event of interest can occur after a competing event.

We will model this situation using a pair of random variables T and D, where T is a non-negative continuous random variable representing the time of an event, and D takes almost surely values $1, \dots, K$, determining the type of event that occurred. Martingale representation will be used, as well as classical survival analysis. Assume that we have time interval $[0, \tau]$, for some $\tau > 0$ finite.

We will work with several fundamental characteristics, including the *cause-specific hazard* of the k-th event. We will also use another type of risk so called the *subdistribution hazard*, presented in the paper Fine and Gray [1999]. Furthermore, we will define the *cumulative incidence function of the k-th event*, which is used in competing risk models.

Let's start by defining the *cause-specific hazard* for the k-th event:

Definition 8 (Cause-specific hazard for the k-th event). *Let T be a non-negative random variable, and random variable D takes values in $1, \dots, K$. The* cause-specific hazard for the k-th event*, for $k \in 1, \dots, K$, is defined by the following expression*

$$\lambda_k^{CS}(t) = \lim_{h \downarrow 0} \frac{P(t \leq T < t + h; D = k | T \geq t)}{h}.$$

Definition 9 (Subdistribution hazard of the k-th event). *Let T be a non-negative random variable, and random variable D takes values in $1, \dots, K$. The* subdistribution hazard for the k-thevent*, for $k \in \{1, \dots, K\}$, is defined by the following expression*

$$\lambda_k^{SD}(t) = \lim_{h \downarrow 0} \frac{P(t \leq T < t+h; D = k | T \geq t \cup (T \leq t \cap D \neq k))}{h}.$$

Definition 10 (Cumulative Incidence Function of the k-th event)**.** *Let T be a non-negative random variable, and let D take values in the range $1, \ldots, K$. The* Cumulative Incidence Function (CIF) *for the k-th event, for $k \in \{1, \ldots, K\}$, is defined by the following expression:*

$$CIF_k(t) = P(T \leq t; D = k).$$

As can be seen from Definitions 8 and 9, subdistribution and cause-specific hazards are very similar, but they differ in terms of which subjects are kept at risk. For cause-specific hazard of event type k, it can be understood that the risk applies to subjects for whom no event occurred. Mathematically, the event $T \geq t$ is used in the condition. This type of hazard essentially extends the hazard definition from classic survival analysis (Definition 2).

The subdistribution hazard for the k-th event leaves subjects at risk if no event has occurred until time t, just like the cause-specific hazard. Additionally, it keeps subjects at risk who have had an event occur before time t, and if it was a different event type than k. Therefore, subjects for whom a different event than the k-th event occurred are kept at risk until infinity. Thus, we can model the impact of explanatory variables on cause-specific and subdistribution hazards using regression (a suitable generalisation of the Cox proportional hazard model). Models that use subdistribution hazard to characterise competing events are sometimes called CIF models, as there is a direct connection to the cumulative incidence function [Fine and Gray, 1999].

In the article Lau et al. [2009], it is argued that using one or the other type of hazard is suitable for different purposes. Cause-specific hazard is used to determine the causes of specific events, in other words, to answer etiological questions. On the other hand, subdistribution hazard is used for prediction. For example, in a medical context, there may be two types of death: heart failure and another type of death. If we want to predict whether a patient will die from heart failure or a different kind of death, it is more appropriate to use subdistribution hazard. We will sometimes refer to SD hazard as a subdistribution hazard and CS hazard as a cause-specific hazard for brevity.

In addition to making inferences about the k-th event, it may be helpful to know the probability that the event will occur after time t (survival function). The survival function will have the same notation as in classical survival analysis, $S(t)$. This characteristic is related to CS hazard. Furthermore, the CIF of the k-th event can also be calculated from the CS hazard functions.

Definition 11 (Cumulative CS hazard of the k-th event)**.** *Let T be a non-negative random variable, and D almost surely takes values in the range $1, \ldots, K$. The* cumulative cause-specific hazard (CS) of the k-th event*, for $k \in \{1, \ldots, K\}$, is defined by the following expression:*

$$\Lambda_k^{CS}(t) = \int_0^t \lambda_k^{CS}(s) ds$$

If we combine all types of events into one, we find ourselves in classical survival analysis. In this case, we will denote $\Lambda(t) = \sum_{j=1}^{K} \Lambda_j^{CS}(t)$ as the overall cumulative hazard function.

Definition 12 (Overall Survival Function)**.** *Let T be a non-negative random variable, and D almost surely takes values in the range $1, \ldots, K$. The* overall survival function *is defined by the following expression:*

$$S(t) = P(T > t).$$

Lemma 4. *If the assumptions for competing risks are met, then the following holds:*

$$S(t) = \exp\left(-\sum_{k=1}^{K} \Lambda_k^{CS}(t)\right).$$

Proof. First, let us introduce auxiliary functions for brevity. Define the function

$$f_k(t) = \lim_{h \downarrow 0} \frac{P(t \leq T < t+h; D = k)}{h}.$$

After introducing this function, we can realise that

$$\lambda_k^{CS}(t) = \lim_{h \downarrow 0} \frac{P(t \leq T < t+h; D = k)}{hS(t-)} = \frac{f_k(t)}{S(t)}.$$

We sum the CS hazards across all event types and integrate both sides of the equation over the interval $(0, t)$. We assume the continuity of T, so $S(t-) = S(t)$.

$$\int_0^t \sum_{k=1}^{K} \lambda_k^{CS}(s) ds = \int_0^t \frac{\sum_{k=1}^{K} f_k(s)}{S(s)} ds + C.$$

Furthermore, we can easily realise that the numerator in the integrand is the derivative of the denominator

$$\frac{d \log(S(u))}{du} = \frac{d \log(\sum_{k=1}^{K} P(T > u; D = k))}{du} = -\frac{\sum_{k=1}^{K} f_k(u)}{S(u)}.$$

Therefore, we obtain that $\sum_{k=1}^{K} \Lambda_k^{CS}(t) = -\log(S(t)) + C$. Substituting $t = 0$, we get $C = 0$, from which the lemma follows. □

Lemma 5. *If the assumptions for competing risks are met, then the following holds:*

$$CIF_k(t) = \int_0^t \lambda_k^{CS}(s) S(s) ds$$

Proof. We proceed similarly to the proof of Lemma 4.

$$\frac{dP(T \leq t; D = k)}{dt} = f_k(t) = \frac{f_k(t) S(t-)}{S(t-)} = \lambda_k^{CS}(t) S(t-)$$

From this, it can be seen that the lemma's wording is obtained by integrating both sides of the equation. □

If we want to determine the overall survival function from Definition 12, we can further proceed in two different ways. The first approach is to combine the event of interest with competing events. We transform the case into classical survival analysis and use the Kaplan-Meier estimate to obtain a consistent survival function estimate. However, if we already have computed estimates of CS hazard or cumulative CS hazard for all events, performing the procedure mentioned above is unnecessary. Using Lemma 4 is sufficient, and we get another possible way to estimate the overall survival function.

2. Nonparametric Estimation

In this chapter, we introduce two approaches. The first approach involves estimating the complement to a survival function ("distribution function") for the event of interest, with the presence of competing events in the data, but this fact is ignored. The second approach takes into account the occurrence of competing risks. We estimate the cumulative incidence function for the event of interest. We always denote the first type of event as an event of interest. Other types of events $\{2, \dots, K\}$ are always competing events; sometimes, we merge competing events just to an event number 2. Asymptotic normality of the estimator of the cumulative incidence function is derived at the end of the chapter.

Let's consider a situation where we have a continuous random variable T and a variable D determining the type of event. Additionally, we have a censoring variable C. Assume that $(T_1, D_1)^T, \dots, (T_n, D_n)^T$ are iid random vectors and we have censoring iid random variables $C_1, \dots, C_n$. These variables (not necessarily all of them) are observed for n individuals, more specifically, for the i-th individual, we observe either the time and type of event (variables T_i and D_i), or the censoring time (C_i). We denote X_i as $min(T_i, C_i)$. To sum up we can see set of independent random vectors

$$(X_1, \mathbf{1}\{T_1 < C_1\}D_1)^T, \dots, (X_n, \mathbf{1}\{T_n < C_n\}D_n)^T,$$

where the second term in each vector is an indicator, which type of event was observed (0 for censoring, k for k-th type of event). Let's have realisations of random variables X_i, without loss of generality sorted as $0 < x_1 < \dots < x_n$.

To present estimators, it is useful to set martingale and counting process notation; it will be easier to derive properties afterwards.

Definition 13 (Counting Process of the j-th Event)**.** *The* counting process of the j-th event for the i-th subject *is defined by the following expression:*

$$N_{i,j}(t) = \mathbf{1}\{T_i \leq t, D_i = j\}$$

The sum over all subjects is denoted as $\overline{N}_j(t) = \sum_{i=1}^n N_{i,j}(t)$.

Note that the counting process is observable. When j-th event is observed, it is not censoring time, and we are observing T_i directly.

Definition 14 (At-risk Process)**.** *The* at risk process of the i-th subject *for the cause-specific risk is defined as:*

$$Y_i(t) = \mathbf{1}\{X_i \geq t\}$$

The sum over all subjects is denoted as $\overline{Y}(t) = \sum_{l=1}^n Y_i(t)$.

2.1 Ignoring Competing Events

The foundation of statistical analysis is to realise that other competing events may exist in the data. When conducting an analysis without knowledge (ignoring) of competing events, the Kaplan-Meier estimate of the survival function is typically used (we will denote it as $\widehat{KM}_1$). The person conducting this analysis thinks that $1 - \widehat{KM}_1(t)$ is an estimator of probability that the event happens until time t. This is actually a cumulative incidence function, $P(T \leq t, D = 1)$. However, competing events are treated the same way as if they were censored in the estimate. A warning against this approach was already mentioned in Section 1.2. The estimator's bias and explanation will be conducted in the following sections. In the established notation, the "incorrect" estimator of the cumulative incidence function is given by:

$$1 - \widehat{KM}_1(t) = 1 - \prod_{s \leq t} \left(1 - \frac{\Delta \overline{N}_1(s)}{\overline{Y}(s)}\right).$$

Additionally, a question arises about what $\widehat{KM}_1$ actually estimates. Putter et al. [2007] suggested that $1 - \widehat{KM}_1$ estimates the following expression:

$$\int_0^t \lambda_1^{CS}(s) \exp\left(-\Lambda_1(s)\right) ds.$$

From this equation and Lemmas 4 and 5, we can deduce the errors or biases that would arise if the presence of competing risks were ignored. In the following inequality is used fact that $\Lambda_j(s) \geq 0$.

$$\int_0^t \lambda_1^{CS}(s) \exp\left(-\sum_{j=1}^{K} \Lambda_j(s)\right) ds \leq \int_0^t \lambda_1^{CS}(s) \exp\left(-\Lambda_1(s)\right) ds.$$

In total, this would lead to an overestimation of the estimate of the cumulative incidence function for the event of interest.

2.2 Taking competing events into account

The cumulative incidence function can be easily interpreted as the probability of event k occurring by time t, so it is useful to present an estimator of this function. To construct such an estimator, we will use Lemma 4. In the lemma, we replace unknown quantities with their estimates. We replace the integral with a sum over all observed times x_i up to time t. We substitute the CS hazard for event k at time x_i with its empirical estimate. We replace the overall survival function with the Kaplan-Meier estimate (left-continuous), combining the event of interest and competing events. The Kaplan-Meier estimate, which combines events of interest and competing events is denoted as:

$$\widehat{KM}(t) = \prod_{s \leq t} \left(1 - \frac{\sum_{j=1}^{K} \Delta \overline{N}_j(s)}{\overline{Y}(s)}\right).$$

Perhaps this notation could be confusing because we commonly use it in classical survival analysis. However, this estimate is a standard Kaplan-Meier estimate in basic survival analysis. Following this procedure, we obtain an estimate of $\widehat{CIF}_k(t)$, which is formally defined.

Definition 15 (Cumulative Incidence Function Estimator)**.** *Under the assumptions for the case of the competing risk, the estimate of the cumulative distribution function for the $j-th$ type of an event is*

$$\widehat{CIF}_j(t) = \sum_{i;x_i \le t} \frac{\Delta \overline{N}_j(x_i)}{\overline{Y}(x_i)} \widehat{KM}(x_i-).$$

Corollary. The estimated cumulative incidence function, as stated in Definition 15, can be rewritten:

$$\int_0^t \frac{\widehat{KM}(u-) d\overline{N}_j(u)}{\overline{Y}(u)}.$$

The next section will indicate how this estimate deals with the occurrence of competing events. With some additional assumptions, it could be proved that the estimate from Definition 15 is consistent for the cumulative incidence function.

2.3 Comparison of Estimates

It was shown that estimator $1 - \widehat{KM_1}(t)$ in Section 2.1 estimates the function which is greater than the cumulative incidence function. There could raise a question if the inequality holds between estimators $1 - \widehat{KM_1}$ and $\widehat{CIF_1}$. Without loss of generality, assume that we are dealing with just two types of events: 1st, an event of interest and 2nd, a competing event. The following lemma will help to prove the statement of inequality.

Lemma 6. *For the problem of competing events, we can rewrite the estimate from Section 2.1 as*

$$1 - \widehat{KM_1}(t) = \int_0^t \widehat{KM_1}(u-) \frac{d\overline{N_1}(u)}{\overline{Y}(u)}.$$

Proof. Function $\widehat{KM_1}(t)$ is a stepwise decreasing function so that it can be expressed as the sum of the initial value (at time $t = 0$) and its steps

$$\widehat{KM_1}(t) = \widehat{KM_1}(0) + \sum_{i:x_i \le t} \Delta \widehat{KM_1}(x_i).$$

By simple calculation, we obtain

$$\Delta \widehat{KM_1}(t) = \widehat{KM_1}(t) - \widehat{KM_1}(t-) = -\widehat{KM_1}(t-) \frac{\Delta \overline{N_1}(t)}{\overline{Y}(t)}.$$

Finally, we have an equation from which the assertion of the lemma could be seen

$$1 - \widehat{KM_1}(t) = \sum_{i:x_i \le t} \widehat{KM_1}(x_i-) \frac{\Delta \overline{N_1}(x_i)}{\overline{Y}(x_i)}.$$

□

Theorem 7 (Inequality of estimators)**.** *In the setting of competing risks, following inequality between two estimators holds*

$$\widehat{CIF}_1(t) \leq 1 - \widehat{KM}_1(t).$$

Proof. We start with an estimator of the overall survival function and with $\widehat{KM}_1$. With an assumption of continuous time we obtain

$$\frac{\widehat{KM}(u-)}{\widehat{KM}_1(u-)} = \frac{\prod_{i;x_i<u}\left(1 - \frac{\Delta\overline{N}_1(x_i)+\Delta\overline{N}_2(x_i)}{\overline{Y}(x_i)}\right)}{\prod_{i;x_i<u}\left(1 - \frac{\Delta\overline{N}_1(x_i)}{\overline{Y}(x_i)}\right)} =$$

$$= \frac{\prod_{i;x_i<u}\left(1 - \frac{1}{\overline{Y}(x_i)}\right)^{\Delta\overline{N}_1(x_i)+\Delta\overline{N}_2(x_i)}}{\prod_{i;x_i<u}\left(1 - \frac{1}{\overline{Y}(x_i)}\right)^{\Delta\overline{N}_1(x_i)}} = \prod_{i;x_i<u}\left(1 - \frac{1}{\overline{Y}(x_i)}\right)^{\Delta\overline{N}_2(x_i)} \leq 1.$$

Continuity of time is a key assumption because there cannot be more events at one time. Now we can easily see an inequality $\widehat{KM}(u-) \leq \widehat{KM}_1(u-)$ and continue with integrating both sides. By the last step, we obtain assertion of the lemma.

$$\widehat{CIF}_1(t) = \int_0^t \widehat{KM}(u-)\frac{d\overline{N_1}(u)}{\overline{Y}(u)} \leq \int_0^t \widehat{KM}_1(u-)\frac{d\overline{N_1}(u)}{\overline{Y}(u)} = 1 - \widehat{KM}_1(t).$$

□

In the article by Gooley et al. [1999], the authors illustrate why the approach presented in Section 2.2 is correct in contrast to the approach in Section 2.1. The estimate of the cumulative incidence function from Section 2.2 can be expressed as a step function, with each step occurring when an event of interest takes place. For the purpose of abbreviation and simplification of notation, we use the following notation. Indicator of an event of interest at the time x_i is denoted by $e_i = \Delta\overline{N_1}(x_i)$. Similarly, the competing event indicator is $r_i = \Delta\overline{N_2}(x_i)$ and indicator of censoring $c_i = 1 - \Delta\overline{N_1}(x_i) - \Delta\overline{N_2}(x_i)$. The number of patients who are known to be at risk of failure beyond time x_i is denoted as $n_i = \overline{Y}(x_{i+1})$

Let's denote the size of the step as J_c for $\widehat{CIF}_1$. Then the estimator can be rewritten as:

$$\widehat{CIF}_1(t) = \sum_{i;x_i\leq t} e_i J_c(x_i),$$

where the function expressing the size of the step has the property stated by Gooley et al. [1999]:

$$J_c(x_i) = J_c(x_{i-1})\left(1 + \frac{c_i}{n_i}\right), \tag{2.1}$$

for $i \in \{2, \ldots, n\}$. Where $J_c(x_1)$ is equal to $1/n$. These step sizes can be imagined as weights being redistributed according to whether censoring occurred, similar to what was discussed in Section 1.1.

Now, let's assume a hypothetical situation where censoring occurred at time x_1 and an event of interest happened at time x_2. According to Equation (2.1), we would have $\widehat{CIF}(x_2) = 1/n$. However, based on Definition 15, it should be

$\widehat{CIF}(x_2) = 1/(n-1)$. From this situation, we can see that Equation (2.1) is incorrect. The following calculation shows how to adjust the recurrence equation to be entirely in line with the definition.

Calculations are based on Proposition 3 and its corollary. We can see from an estimator of the cumulative incidence function

$$\widehat{CIF}_1(t) = \sum_{i;x_i \leq t} \Delta\overline{N_1}(x_i) \frac{\widehat{KM}(x_i-)}{\overline{Y}(x_i)},$$

that $J_c(x_i) = \widehat{KM}(x_i-)/\overline{Y}(x_i)$.

We can use Proposition 3. As a result, we obtain

$$J_c(x_i) = \frac{1}{n\prod_{j<i}\left(1-\frac{1}{n_{j-1}}\right)^{c_j}} = \frac{1}{n}\prod_{j<i}\left(\frac{n_{j-1}}{n_{j-1}-1}\right)^{c_j} =$$
$$= \frac{1}{n}\prod_{j<i}\left(1+\frac{1}{n_{j-1}-1}\right)^{c_j} = J_c(x_{i-1})\left(1+\frac{c_{i-1}}{n_{i-1}}\right).$$

The last expression states correction to Equation (2.1) proposed by Gooley et al. [1999].

Similarly, the naive (incorrect) incidence function estimate determined by the formula $1 - \widehat{KM}_1(t)$ is also a step function where the jumps occur precisely at event times. Thus similarly it can be expressed in the form of a sum:

$$1 - \widehat{KM}_1(t) = \sum_{i;x_i \leq t} e_i J_{km}(x_i).$$

In this case, the jump function has the following property according to Gooley et al. [1999]:

$$J_{km}(x_i) = J_{km}(x_{i-1})\left(1+\frac{c_i + r_i}{n_i}\right). \tag{2.2}$$

This property shows that this estimate treats competing events as censoring from the idea of reweighting presented in Section 1.1. However, once again, a calculation is provided to demonstrate that the equality 2.2 is incorrect.

According to Lemma 6, estimate could be written as a sum

$$1 - \widehat{KM_1}(t) = \sum_{x_i \leq t} \widehat{KM_1}(x_i-) \frac{\Delta\overline{N_1}(x_i)}{\overline{Y}(x_i)}.$$

From this expression, it follows that

$$J_{km}(x_i) = \frac{\widehat{KM_1}(x_i-)}{\overline{Y}(x_i)}.$$

Now we use Proposition 3 and its corollary. We express the desired recurrent equation with algebraical operations and a switch to the new stated notation.

$$
\begin{aligned}
J_{km}(x_i) &= \frac{\widehat{KM}_1(x_i-)}{n\widehat{KM}(x_i-)\hat{G}(x_i-)} \\
&= \frac{\prod_{j<i}\left(1-\frac{1}{n_{j-1})}\right)^{e_j}}{n\prod_{j<i}\left(1-\frac{1}{n_{j-1}}\right)^{e_j+r_j}\prod_{j<i}\left(1-\frac{1}{n_{j-1}}\right)^{c_j}} \\
&= \frac{1}{n\prod_{j<i}\left(1-\frac{1}{n_{j-1}}\right)^{c_j+r_j}} = \frac{\prod_{j<i}\left(1+\frac{1}{n_{j-1}-1}\right)^{c_j+r_j}}{n} \\
&= J_{km}(x_{i-1})\left(1+\frac{c_{i-1}+r_{i-1}}{n_{i-1}}\right)
\end{aligned}
$$

The calculation shows a similar equality to the original one but with index shifts.

2.4 Asymptotic Properties of the Estimator

When deriving the asymptotic distribution of the cumulative incidence function estimator, the cumulative cause-specific hazard properties are utilized. This can be estimated analogously to the Nelson-Aalen estimator of cumulative hazard in survival analysis. The estimation of cumulative cause-specific hazard, and subsequently the cumulative incidence of the first event, can be expressed by using counting processes and martingales. Now, let's introduce key definitions necessary for constructing the filtration and determining the subsequent martingale.

Definition 16 (Nelson-Aalen Estimator of Cumulative Cause-Specific hazard)**.** *The* Nelson-Aalen estimator *of cumulative cause-specific hazard for the j-th event is defined by the following expression:*

$$
\hat{\Lambda}_j^{CS}(t) = \int_0^t \frac{d\overline{N}_j(u)}{\overline{Y}(u)}.
$$

Corollary. One could observe that Nelson-Aalen estimate of overall cumulative hazard is a sum of estimates of cumulative cause-specific hazards:

$$
\hat{\Lambda}(t) = \sum_{j=1}^{K} \hat{\Lambda}_j^{CS}(t).
$$

For the martingale definition, filtration is needed, so we denote the filtration for the j-th type of event for all of the data by

$$
\mathcal{F}_t^j = \sigma\{N_{i,j}(s), Y_i(s), 0 \leq s \leq t, i = 1, \dots, n\}.
$$

Theorem 8 (Martingale for Cause-Specific hazard)**.** *Let λ_j^{CS} be the CS hazard of j-th event , then*

$$
M_{i,j} = N_{i,j}(t) - \int_0^t Y_i(u)\lambda_j^{CS}(u)du
$$

is a martingale with respect to $\sigma(N_{i,j}(s), Y_i(s), 0 \leq s \leq t)$ if and only if the following equality holds for each t

$$\lambda_j^{CS}(t) = \lim_{h \downarrow 0} \frac{P(t \leq T_i < t+h; D_i = j \mid X_i \geq t)}{h}. \tag{2.3}$$

Proof. The proof is just an analogous application of the proof of Theorem 2. For more details see Andersen et al. [2012, part IV.1] □

Remark. Typically, we aim to measure the risk of only one competing event, while we are no longer interested in the rest. Therefore, we switch to the notation where $D = 1$ represents the event of interest and $D = 2$ represents competing events. Thus, we have combined all types of competing events.

The asymptotic properties of cause-specific hazards and cumulative incidence function are derived using martingale theory. This is the reason for mentioning theorems like Theorem 8. Just before we introduce a theorem about asymptotic properties of the estimator, we state and remind theorem about cumulative cause-specific hazard and overall cumulative hazard.

Theorem 9. *Let the random vectors $(T_1, D_1, C_1)^T, \ldots, (T_n, D_n, C_n)^T$ are independently distributed. Suppose variables $C_1, \ldots, C_n$ to be identically distributed as well suppose that for vectors $(T_1, D_1)^T, \ldots, (T_n, D_n)^T$. Assume that there exist $\gamma > 0$ such that $P(Y_i(\tau) = 1) > \gamma > 0$. Let's assume the condition of independent censoring (2.3) for each $j \in \{1, \ldots, K\}$. It holds:*

1. $\sqrt{n}\left(\widehat{\Lambda}_j^{CS}(t) - \Lambda_j^{CS}(t)\right) = \sqrt{n}\int_0^t \frac{\mathbf{1}\{\overline{Y}(u)>0\}d\overline{M}_j(u)}{\overline{Y}(u)} + o_p(1)$
2. $\sqrt{n}\left(\widehat{\Lambda}(t) - \Lambda(t)\right) = \sqrt{n}\int_0^t \frac{\mathbf{1}\{\overline{Y}(u)>0\}d\overline{M}(u)}{\overline{Y}(u)} + o_p(1)$

Proof. Proofs can be found in Andersen et al. [2012, part IV.1] □

Cumulative incidence functions can be suitably expressed in a similar way as in the previous theorem. From this representation, it is possible to utilize the central limit theorem for martingales. But let's state two more lemmas that will be useful afterwards.

Lemma 10. *Assume the condition of independent censoring, then for every $t \geq 0$ such that $S(t) > 0$, it holds*

$$\frac{\widehat{KM}(t)}{S(t)} - 1 = -\int_0^t \frac{\widehat{KM}(u-)}{S(u)} d\left(\widehat{\Lambda} - \Lambda\right)(u).$$

Proof. The proof of this lemma is just technical. It is an exercise from the theory of measure, especially per partes for Lebesgue-Stieltjes integral. □

Lemma 11. *Let $A_n(t)$ is a random process that approaches weakly in $D[0,\tau]$ to zero process. Then $A_n(t)$ converges uniformly in probability to* 0.

Proof. Let $A \in D[0,\tau]$. Transformation $A \longrightarrow \sup_{t\in[0,\tau]} |A|$ is continuous with respect to metric of space $D[0,\tau]$. Assume that $A_n \xrightarrow{D[0,\tau]} A$. With the usage of the continuous mapping theorem, we obtain convergence in distribution

$$\sup_{t\in[0,\tau]} |A_n| \xrightarrow{D} \sup_{t\in[0,\tau]} |A|.$$

As a result, the zero mean process has a supremum equal to zero; this concludes the proof because convergence in distribution to constant is equivalent to convergence in probability. □

Theorem 12 (Martingale Representation of $\widehat{CIF}_j$)**.** *Assuming the conditions of Theorem 9 and $S(t) > 0$, the following holds:*

$$\sqrt{n}\left(\widehat{CIF}_j(t) - CIF_j(t)\right) = \sqrt{n}\left(\int_0^t \widehat{KM}(u-)d\hat{\Lambda}_j^{CS}(u) - \int_0^t S(u-)d\Lambda_j^{CS}(u)\right)$$

$$=\sqrt{n}\int_0^t \frac{S(u-)\mathbf{1}\{\overline{Y}(u) > 0\}d\overline{M}_j(u)}{\overline{Y}(u)} + \sqrt{n}CIF_j(t)\int_0^t \frac{\mathbf{1}\{\overline{Y}(u) > 0\}d\overline{M}(u)}{\overline{Y}(u)}$$

$$-\sqrt{n}\int_0^t \frac{CIF_j(u)\mathbf{1}\{\overline{Y}(u) > 0\}d\overline{M}(u)}{\overline{Y}(u)} + o_p(1)$$

The expression $o_p(1)$ means that this part of the formula converges in probability to zero as n tends to infinity. Furthermore, the convergence is uniform.

Proof. A derivation of the first part of the theorem follows the paper published by Lin [1997]. Some important parts of the proof in the paper are missing. We have completed these steps and proved some parts in a different way.

It is easy to see that $\sqrt{n}\left(\widehat{CIF}_j(t) - CIF_j(t)\right)$ is equal to

$$\sqrt{n}\int_0^t \widehat{KM}(u-)d\left(\hat{\Lambda}_j^{CS} - \Lambda_j^{CS}\right)(u) + \sqrt{n}\int_0^t \left(\widehat{KM}(u-) - S(u)\right)d\Lambda_j^{CS}(u). \quad (2.4)$$

The first term of the sum in the Equation (2.4) is equivalent to

$$\sqrt{n}\int_0^t S(u-)d\left(\hat{\Lambda}_j^{CS} - \Lambda_j^{CS}\right)(u) + o_p(1).$$

To prove this fact we use Theorem 9. Then we apply the central limit theorem. See Theorem 18 in the appendix. We prove that the expression converges weakly to a trivial zero process, but first we have to prove the assumptions of the central limit theorem.

$$\sqrt{n}\left(\int_0^t S(u-) - \widehat{KM}(u-)d\left(\hat{\Lambda}_j^{CS} - \Lambda_j^{CS}\right)(u)\right) =$$

$$= \sqrt{n}\left(\int_0^t \left(S(u-) - \widehat{KM}(u-)\right)\frac{\mathbf{1}\{\overline{Y}(u) > 0\}d\overline{M}_j(u)}{\overline{Y}(u)}\right) + o_p(1) =$$

$$= \sum_{i=1}^n \sqrt{n}\int_0^t \left(S(u-) - \widehat{KM}(u-)\right)\frac{\mathbf{1}\{\overline{Y}(u) > 0\}dM_{i,j}(u)}{\overline{Y}(u)}$$

$$:= U_j^{(n)}(t) = \sum_{i=1}^n U_{i,j}^{(n)}$$

It is easy to see that the integrand is a bounded and left-continuous function. As a consequence of left continuity, it is a predictable function. Let's calculate predictable variation

$$\langle U_j^{(n)}, U_j^{(n)}\rangle(t) = \int_0^t \left(S(u-) - \widehat{KM}(u-)\right)^2 \frac{\mathbf{1}\{\overline{Y}(u) > 0\}d\Lambda_j^{CS}(u)}{\overline{Y}(u)/n}$$

$$\leq \left(\sup_{[0\leq u\leq\tau]}\left(S(u-) - \widehat{KM}(u-)\right)\right)^2 \int_0^t \frac{\mathbf{1}\{\overline{Y}(u) > 0\}d\Lambda_j^{CS}(u)}{\overline{Y}(u)/n}.$$

From the last expression it can be seen that the supremum tends to 0, since the Kaplan-Meier estimate is uniformly consistent. The second part converges in probability to $\int_0^t \lambda_j^{CS}(u)/P(Y_i(u)=1)du$. To sum up, predictable variance converges to 0 in probability. Therefore, it is not necessary to verify Feller-Linderberg's assumption. Finally, we got that the sum of martingale differences converges weakly to zero mean and zero variance Gaussian process so that the process is 0 almost surely. With the usage of the Lemma 11, we have the desired result. We denote $P(Y_i(u)=1) := \pi(u)$.

Next, we modify Equation (2.4). We can see that the second expression of the sum in the equation is equal to $\sqrt{n}\int_0^t \widehat{KM}(u) - S(u)d\Lambda_j^{CS}(u) + o_p(1)$. The reason why this can be done is based on the fact that difference $\widehat{KM}(u) - \widehat{KM}(u-)$ is equal to $\widehat{KM}(u-)/\overline{Y}(u)$ in the continuous case. As a result, we obtain an equation

$$\frac{1}{\sqrt{n}}\int_0^t \widehat{KM}(u-)\frac{d\Lambda_j^{CS}(u)}{\overline{Y}(u)/n}. \tag{2.5}$$

The basic key to prove this statement is that Equation (2.5) consists of part which is bounded in probability and $1/\sqrt{n}$ which together converges to zero. It is not presented a detailed proof since it is just technical.

Now we go on with rewriting the second term in Equation (2.4)

$$\sqrt{n}\int_0^t \left(\widehat{KM}(u) - S(u)\right) d\Lambda_j^{CS}(u) + o_p(1) =$$
$$\sqrt{n}\int_0^t \left(\frac{\widehat{KM}(u)}{S(u)} - 1\right) S(u)d\Lambda_j^{CS}(u) + o_p(1).$$

Based on the result from Lemma 5 we have $S(u)d\Lambda_j^{CS}(u) = dCIF_j(u)$. Furthermore we use Lemma 10. We obtain double integral and again we rewrite it to more convenient expression

$$\sqrt{n}\int_0^t -\int_0^u \left(\frac{\widehat{KM}(s-)}{S(s)}\right) d\left(\widehat{\Lambda} - \Lambda\right)(s)dCIF_j(u) + o_p(1) =$$
$$= \sqrt{n}\int_0^t \int_0^u \left(\frac{\widehat{KM}(s-) - S(s)}{S(s)} + 1\right) d\left(\widehat{\Lambda} - \Lambda\right)(s)dCIF_j(u) + o_p(1) =$$
$$= \int_0^t o_p(1)dCIF_j(u) + \sqrt{n}\int_0^t \left(\widehat{\Lambda}(u) - \Lambda(u)\right) dCIF_j(u) + o_p(1).$$

The last formula consists of the final desired result and integral negligible in probability, again the negligibility does not depend on t. It can be shown again by uniform consistency of a Kaplan-Meier estimate and by similar usage of the central limit theorem for the sum of martingale differences as was provided before. By the last discussion was proven that the Equation (2.4) is equal to

$$\sqrt{n}\int_0^t S(u-)d\left(\widehat{\Lambda}_j^{CS} - \Lambda_j^{CS}\right)(u) + \sqrt{n}\int_0^t \left(\widehat{\Lambda}(u) - \Lambda(u)\right) dCIF_j(u) + o_p(1) \tag{2.6}$$

As the next step, we use integrating by parts for Lebesgue-Stieltjes on the second part of the Equation (2.6). We got

$$\sqrt{n}\int_0^t S(u-)d\left(\widehat{\Lambda}_j^{CS} - \Lambda_j^{CS}\right)(u) + \sqrt{n}\left(\widehat{\Lambda}(t) - \Lambda(t)\right)CIF_j(t)$$
$$-\sqrt{n}\int_0^t CIF_j(u)d\left(\widehat{\Lambda} - \Lambda\right)(u) + o_p(1).$$

To finalise the proof we need to plugin formulas for

$$\sqrt{n}\left(\widehat{\Lambda}(t)-\Lambda(t)\right) \text{ and } \sqrt{n}\left(\widehat{\Lambda}_j(t)-\ \Lambda_j(t)\right),$$

which are provided in Theorem 9. We receive the final formula which is an assertion of the theorem. Note that the term $o_p(1)$ tends to zero uniformly. We obtained final expression

$$\begin{aligned}&\sqrt{n}\int_0^t \frac{S(u-)\mathbf{1}\{\overline{Y}(u)>0\}d\overline{M}_j(u)}{\overline{Y}(u)}\\&+\sqrt{n}CIF_j(t)\int_0^t \frac{\mathbf{1}\{\overline{Y}(u)>0\}d\overline{M}(u)}{\overline{Y}(u)}\\&-\sqrt{n}\int_0^t \frac{CIF_j(u)\mathbf{1}\{\overline{Y}(u)>0\}d\overline{M}(u)}{\overline{Y}(u)}+o_p(1).\end{aligned}$$

□

Remark. As a special result from Theorem 12 we could obtain an asymptotic representation for an event of interest ($j=1$). If we merge all competing events into one, there is a possibility for just two values of j either equal to one or two. In this specific setting of problem, it could be easily proved just by algebraic operations and usage of $S(t)=1-CIF_1(t)-CIF_2(t)$ that

$$\begin{aligned}\sqrt{n}\Big[\int_0^t \frac{(1-CIF_2(u))\,\mathbf{1}\{\overline{Y}(u)>0\}d\overline{M}_1(u)}{\overline{Y}(u)}+\int_0^t \frac{CIF_1(u)\mathbf{1}\{\overline{Y}(u)>0\}d\overline{M}_2(u)}{\overline{Y}^2(u)}\\-CIF_1(t)\int_0^t \frac{\mathbf{1}\{\overline{Y}(u)>0\}(d\overline{M}_2(u)+d\overline{M}_1(u))}{\overline{Y}(u)}\Big]+o_p(1).\end{aligned}$$

Remark. There is used notation $S(u-)$ for left continuous overall survival function. Our assumption is that time to event is a continuous random variable, so it holds $S(t-)=S(t)$. We just wanted to emphasise the fact that integrand is left continuous and as a consequence predictable.

Theorem 13 (Asymptotic normality of $\widehat{CIF_1}$)**.** *Assuming the conditions of Theorem 9.*

The expression

$$\sqrt{n}\left(\widehat{CIF}_1(t)-CIF_1(t)\right)$$

weakly converges to a Gaussian process with zero mean and covariance function $\psi(s,t)$ *for* $t\leq s$

$$\int_0^t \frac{\left(1-CIF_2(u)-CIF_1(t)\right)^2\lambda_1^{CS}(u)}{\pi(u)}du+\int_0^t \frac{\left(CIF_1(u)-CIF_1(t)\right)^2\lambda_2^{CS}(u)}{\pi(u)}du.$$

Proof. The theorem can be proved by two key steps. Firstly, central limit theorem for martingale differences is applied, Theorem 18. As the second key step, we use functional delta limit theorem to prove the assertion of the theorem.
Let's verify the assumptions. First, continuous time is assumed, so we do not

observe two types of events at one specific time. Set notation in the context of the central limit theorem

$$U_1^{(n)}(t) = \sum_{i=1}^{n} U_{i,1}^{(n)} = \sum_{i=1}^{n} \int_0^t \sqrt{n}\frac{(1 - CIF_2(u) - CIF_1(t))\,\mathbf{1}\{\overline{Y}(u) > 0\}}{\overline{Y}(u)} dM_{i,1}(u),$$

$$U_2^{(n)}(t) = \sum_{i=1}^{n} U_{i,2}^{(n)} = \sum_{i=1}^{n} \int_0^t \sqrt{n}\frac{(CIF_1(u) - CIF_1(t))\,\mathbf{1}\{\overline{Y}(u) > 0\}}{\overline{Y}(u)} dM_{i,2}(u).$$

The processes in integrals are denoted as $\sqrt{n}H_1(u)$ and $\sqrt{n}H_2(u)$. We have multivariate counting process of i-th individual $\boldsymbol{N}_i = (N_{i,1}, N_{i,2})^T$. As a consequence we have

$$\begin{pmatrix} M_{i,1} \\ M_{i,2} \end{pmatrix} = \begin{pmatrix} N_{i,1} \\ N_{i,2} \end{pmatrix} - \begin{pmatrix} \int_0^t Y_i(u)\lambda_1^{CS}(u)du \\ \int_0^t Y_i(u)\lambda_2^{CS}(u)du \end{pmatrix}.$$

We can see that both $M_{i,1}$ and $M_{i,2}$ are martingales with respect to filtration $\mathcal{F}^{12} = \sigma\{\boldsymbol{N}_i(s), Y_i(s), 0 \le s \le t, i = 1, \ldots, n\}$. It can be seen that both processes in the integrals are bounded and predictable (left-continuity).

Let's now calculate a function to which predictable variation converges

$$\begin{aligned}
&\langle U_1^{(n)}; U_1^{(n)}\rangle(t) = \int_0^t nH_1^2(u)\overline{Y}(u)d\Lambda_1^{CS}(u) \\
&= \int_0^t \frac{(1 - CIF_2(u) - CIF_1(t))^2}{P(Y_i(u) = 1)}\lambda_1^{CS}(u)du \\
&+ \int_0^t (1 - CIF_2(u) - CIF_1(t))^2 \left(\frac{\mathbf{1}\{\overline{Y}(u) > 0\}}{\overline{Y}(u)/n} - \frac{1}{P(Y_i(u) = 1)}\right) d\Lambda_1^{CS}(u) \\
&= \int_0^t \frac{(1 - CIF_2(u) - CIF_1(t))^2}{P(Y_i(u) = 1)}\lambda_1^{CS}(u)du + A_n \\
&= \int_0^t f_1^2(u)du + o_p(1).
\end{aligned}$$

From the last equality, we have to prove that the second term in the sum is negligible in probability. This is straightforward since we can use inequality $(1 - CIF_1(u) - CIF_2(t))^2 \le 3^2$. We obtain

$$|A_n| \le 9 \int_0^t \left| \left(\frac{\mathbf{1}\{\overline{Y}(u) > 0\}}{\overline{Y}(u)/n} - \frac{1}{P(Y_i(u) = 1)}\right) \right| d\Lambda_1^{CS}(u).$$

From the last expression by uniform convergence of $n/\overline{Y}(u)$ to $1/P(Y_i(u))$ and fact that $\Lambda_1^{CS}(\tau) < \infty$, it could be proved that A_n is negligible in probability. It is just a technical proof and it is omitted. Now let's calculate predictable variation for $U_2^{(n)}$

$$
\begin{aligned}
\langle U_2^{(n)}; U_2^{(n)}\rangle(t) &= \int_0^t nH_2^2(u)\overline{Y}(u)d\Lambda_2^{CS}(u)\\
&= \int_0^t \frac{(CIF_1(u) - CIF_1(t))^2}{P(Y_i(u)=1)}\lambda_2^{CS}(u)du\\
&+ \int_0^t (CIF_1(u) - CIF_1(t))^2 \left(\frac{\mathbf{1}\{\overline{Y}(u)>0\}}{\overline{Y}(u)/n} - \frac{1}{P(Y_i(u)=1)}\right) d\Lambda_2^{CS}(u)\\
&= \int_0^t \frac{(CIF_1(u) - CIF_1(t))^2}{P(Y_i(u)=1)}\lambda_2^{CS}(u)du + A_n^{'}\\
&= \int_0^t f_2^2(u)du + o_p(1).
\end{aligned}
$$

In a similar way as previously, it can be proved ($CIF_1(u) - CIF_1(t) \leq 2$) that $A_n^{'}$ is negligible in probability. Now we can move on to verifying Linderberg's assumption to fulfil all assumptions of the central limit theorem. Let $\varepsilon > 0$. For $\overline{Y}(u) = 0$ the integrand is zero. If $\overline{Y}(u) > 0$ then we have

$$
\begin{aligned}
\langle U_{\varepsilon,1}^{(n)}; U_{\varepsilon,1}^{(n)}\rangle(t) &= \int_0^t nH_1^2(u)\mathbf{1}\{\sqrt{n}H_1(u) > \varepsilon\}\overline{Y}(u)d\Lambda_1^{CS}(u)\\
&= \int_0^t \frac{(1 - CIF_2(u) - CIF_1(t))^2}{\overline{Y}(u)/n}\mathbf{1}\{\sqrt{n}H_1(u) > \varepsilon\}d\Lambda_1^{CS}(u)\\
&= \int_0^t \frac{(1 - CIF_2(u) - CIF_1(t))^2}{\pi(u)}\mathbf{1}\{\sqrt{n}H_1(u) > \varepsilon\}d\Lambda_1^{CS}(u) + o_p(1).
\end{aligned}
$$

In the last equation, the same ideas as before were used, namely, uniform convergence of $\overline{Y}(u)/n$. We obtain expression

$$
\begin{aligned}
&\int_0^t \frac{(1 - CIF_2(u) - CIF_1(t))^2}{\pi(u)}\mathbf{1}\{\sqrt{n}H_1(u) > \varepsilon\}d\Lambda_1^{CS}(u)\\
&\leq \int_0^t \frac{3^2}{\pi(u)}\mathbf{1}\left\{\frac{\overline{Y}(u)}{n\,(1 - CIF_2(u) - CIF_1(t))} < \frac{1}{\sqrt{n}\varepsilon}\right\}d\Lambda_1^{CS}(u)\\
&\qquad\leq \frac{3^2}{\pi(u)}\mathbf{1}\left\{\frac{\overline{Y}(\tau)}{n3} < \frac{1}{\sqrt{n}\varepsilon}\right\}\Lambda_1^{CS}(\tau).
\end{aligned}
$$

From the last inequality, we are able to see that the Feller-Lindenberg assumption is satisfied. The fraction $1/(\sqrt{n}\varepsilon)$ converges to zero and $\overline{Y}(u)/(6n)$ converges uniformly in probability to $\pi(u)/6 > 0$. The Feller-Lindeberg assumption for the second predictable variation could be done by the same procedure.

Finally, we verified all of the assumptions, so we can claim that

$$
\left(U_1^{(n)}(t), U_2^{(n)}(t)\right)^T \xrightarrow{D^2[0,\tau]} \left(\int_0^t f_1(u)dW_1(u), \int_0^t f_2(u)dW_2(u)\right)^T,
$$

where W_1, W_2 are independent Brownian motions. Now if we stop in one specific time t_1, we obtain convergence in distribution. By Cramer-Wold theorem we could find out pointwise asymptotic distribution of the estimator of a cumulative incidence function. However assertion of this theorem is stronger. It is necessary to use the functional delta theorem.

Let ϕ is a mapping (sum of two random processes) from $D^2[0,\tau]$ to $D[0,\tau]$. We can realise that Haddamard derivative of ϕ is actually mapping itself. To verify this, we need continuity and linearity. It is clear that mapping is linear. To prove continuity, let $(g_1^n, g_2^n)^T, (g_1, g_2)^T \in D^2[0,\tau]$ such that

$$(g_1^n, g_2^n)^T \xrightarrow{D^2[0,\tau]} (g_1, g_2)^T .$$

In other words

$$\left\| \begin{pmatrix} g_1^n \\ g_2^n \end{pmatrix} - \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} \right\|_{D^2[0,\tau]}^2 = \|g_1^n - g_1\|_{D[0,\tau]}^2 + \|g_2^n - g_2\|_{D[0,\tau]}^2 \xrightarrow{n\to\infty} 0.$$

From this expression we can see that both terms in sum have to converge to 0 as n tends to infinity. Finally, we can prove continuity from the following inequalities

$$\left\| \phi \begin{pmatrix} g_1^n \\ g_2^n \end{pmatrix} - \phi \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} \right\|_{D[0,\tau]}^2 = \|g_1^n - g_1 + g_2^n - g_2\|_{D[0,\tau]}^2$$

$$\leq \|g_1^n - g_1\|_{D[0,\tau]}^2 + \|g_2^n - g_2\|_{D[0,\tau]}^2 + 2\,\|g_1^n - g_1\|_{D[0,\tau]}\,\|g_2^n - g_2\|_{D[0,\tau]} .$$

Let $a_n \to \infty$ is real sequence and h_n, h sequences in $D^2[0,\tau]$ such that $h_n \xrightarrow{D^2[0,\tau]} h$. Let $\theta \in D^2[0,\tau]$, then

$$a_n \left(\phi \left(\theta + \frac{h_n}{a_n} \right) - \phi(\theta) \right) = \frac{a_n}{a_n} \phi(h_n) \xrightarrow{D[0,\tau]} \phi(h).$$

Linearity and continuity properties were used within calculations. As a result, ϕ is Haddamard differentiable (Definition 30) and we can use the functional delta theorem, Theorem 19. We obtain

$$U_1^{(n)}(t) + U_2^{(n)}(t) \xrightarrow{D[0,\tau]} \int_0^t f_1(u)dW_1(u) + \int_0^t f_2(u)dW_2(u).$$

It is easy to see that a sum of two independent, time-transformed Brownian motions is again a time-transformed Brownian motion. As a consequence limiting process of an estimator of a cumulative incidence function is a zero-mean Gaussian process with a variance function $\psi(s,t)$ for $t \leq s$

$$\int_0^t \frac{(1 - CIF_2(u) - CIF_1(t))^2 \lambda_1^{CS}(u)}{\pi(u)} du + \int_0^t \frac{(CIF_1(u) - CIF_1(t))^2 \lambda_2^{CS}(u)}{\pi(u)} du.$$

□

Corollary. If we are able to estimate the variance function in the previous theorem then we are able to perform the point inference with cumulative incidence functions.

Results of Theorem 9 could be weaker and proof could be easier as was already discussed in the proof. However, convergence of the cumulative incidence function stopped at specific times is not enough for the construction of the confidence bands. As a consequence of stronger results of the theorem, we are able to construct such confidence bands. An example of this construction can be seen for instance in Lin [1997].

For pointwise confidence intervals of CIF_1 in practical situation we need to estimate a variance function. The variance function in Theorem 13 could be estimated by

$$\int_0^t \frac{\left(1 - \widehat{CIF_2}(u) - \widehat{CIF_1}(t)\right)^2 d\overline{N_1}(u)}{\overline{Y}^2(u)/n} + \int_0^t \frac{\left(\widehat{CIF_1}(u) - \widehat{CIF_1}(t)\right)^2 d\overline{N_2}(u)}{\overline{Y}^2(u)/n}.$$

The same estimator of variance function is used by R package *cmprsk*, Gray [2022] based on Aalen [1978].

3. Semiparametric Estimation

Similar to survival analysis, sometimes it is necessary to examine the complex effects of multiple variables, potentially continuous, simultaneously on the risk of an event. In survival analysis, the Cox proportional hazards model is the fundamental model for this situation. Similar models can be supposed in the context of competing risks. This chapter will present two approaches: cause-specific and subdistribution hazard.

3.1 Cause-Specific hazard

As already defined in the introductory chapter in Definition 9, CS hazard is one possible extension of a classical hazard. Assume that CS hazard possibly depends on some covariates. Let us have a regression parameter $\boldsymbol{\beta_j}$ (p-dimensional vector) for the j-th type of event and an explanatory variable vector $\boldsymbol{Z}(t)$. Furthermore, denote true value of parameters $\boldsymbol{\beta_j^0}$ and a baseline hazard $\lambda_{0,j}^{CS}$. Assume that the following equation holds

$$\lambda_{\boldsymbol{Z},j}^{CS}(t) = \lambda_{0,j}^{CS}(t) \exp\left(\boldsymbol{Z}^T(t)\boldsymbol{\beta_j^0}\right). \tag{3.1}$$

Based on this assumption, we can construct the partial likelihood as in the Cox model. We can use the martingale notation introduced in Section 2.4. We only add that $\boldsymbol{Z_i}(t)$ is the vector of explanatory variables for the i-th subject. Note that we have to assume that the vectors of processes $\boldsymbol{Z_i}(t)$ are left continuous with right-hand side limits. Let's provide the definition in this notation.

Definition 17. *The function*

$$\mathbf{L}(\boldsymbol{\beta_1}, \ldots, \boldsymbol{\beta_K}) = \prod_{j=1}^{K}\prod_{i=1}^{n}\prod_{s>0}\left(\frac{Y_i(s)\exp\left(\boldsymbol{\beta_j^T Z_i}(s)\right)}{\sum_{l=1}^{n} Y_l(s)\exp\left(\boldsymbol{\beta_j^T Z_l}(s)\right)}\right)^{\Delta N_{i,j}(s)}$$

is called the partial likelihood of cause-specific hazard, *where* $\Delta N_{i,j}(t)$ *denotes the change in the counting process at time* t.

Maximising this function provides estimates for the coefficients $\boldsymbol{\beta_1}, \ldots, \boldsymbol{\beta_K}$. It is noteworthy that according to this definition, we are essentially performing the standard Cox model K times. More specifically, if we consider the part with $j = 1$, this part is essentially a Cox model in which all other events are treated as censoring so that there can be used theory developed for the Cox model in classical survival analysis.

Assume that we have an exposed and unexposed group of patients. We are focused on the effect of exposure on the event of interest by the hazard ratio of CS hazard $HR_1^{CS}(t) = \lambda_{E,1}^{CS}(t)/\lambda_{U,1}^{CS}(t)$. In the same way for CS hazard ratio for the second event $HR_2^{CS}(t) = \lambda_{E,2}^{CS}(t)/\lambda_{U,2}^{CS}(t)$. In the context of the Cox model, the CS hazard ratio is assumed to be constant over time. The fact that HR_1^{CS} is greater than one does not necessarily imply that the cumulative incidence of the first event is greater for the exposed than for the unexposed group. As a consequence, we cannot use it to make predictions. In general CS hazards for all events should be analysed including baseline CS hazards. However, it cannot be said that this approach is unusable. If we estimate the cause-specific hazard of the event of interest and all of the competing events, we can express it through a transformation of the relationship with the cumulative incidence function, which can be appropriately interpreted (see Lemma 4 and 5). Therefore, this approach requires estimating the cause-specific hazards of all types of events. On the other hand, these estimates can be executed using software that can implement the standard Cox model. For more details see Lau et al. [2009] and Allignol et al. [2011]. Since it is hard to predict according to CS hazards, there is motivation for models of SD hazard presented in the upcoming section.

Eventhough we mentioned some issues with the interpretation, it is still a useful approach, therefore we present some important definitions and propositions about the estimators. We can rewrite partial likelihood (just j-th part) into the logarithmic form and corresponding score function, but first, we will present the notation.

Let

$$S_n^{(k)}(\boldsymbol{\beta}_j, t) = \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\boldsymbol{Z_i}^{\bigotimes k}(t)\exp\left(\boldsymbol{\beta_j}^T \boldsymbol{Z_i}(t)\right),$$

where $z^{\bigotimes 0} = 1$, $z^{\bigotimes 1} = z$ and $z^{\bigotimes 2} = zz^T$. Let

$$\overline{\boldsymbol{Z}}_n(\boldsymbol{\beta_j}, t) = \frac{\boldsymbol{S}_n^{(1)}(\boldsymbol{\beta}_j, t)}{S_n^{(0)}(\boldsymbol{\beta}_j, t)}.$$

Definition 18 (Logarithmic partial likelihood for $j-$th CS hazard)**.** *The function*

$$\ell(\boldsymbol{\beta_j}) = \sum_{i=1}^{n}\int_0^\infty \left[\boldsymbol{\beta_j}^T \boldsymbol{Z_i}(s) - log(nS_n^0(\boldsymbol{\beta_j}, s))\right] dN_{i,j}(s).$$

is called logarithmic partial likelihood for $j-th$ cause-specific hazard.

By differentiating of log partial likelihood with respect to parameter $\boldsymbol{\beta_j}$ we obtain the score function.

Definition 19 (Score function of $j-$th CS hazard)**.**

$$\boldsymbol{U}_{n,j}(\boldsymbol{\beta_j}) = \sum_{i=1}^{n}\int_0^\infty \left[\boldsymbol{Z_i}(t) - \overline{\boldsymbol{Z}}_n(\boldsymbol{\beta_j}, t)\right] dN_{i,j}(t)$$

From the last definition, we can see the importance of Theorem 3.1, since we have a sum of integrals which can be transformed into a sum of martingales. The consistency of the estimator of $\boldsymbol{\beta_j}$ can be proved with some additional assumptions. The asymptotic properties in detail are omitted since the procedure is the same as in ordinary survival analysis.

3.2 Subdistribution hazard

Semiparametric models using subdistribution hazard are sometimes referred to as CIF models, Austin et al. [2016], as there is a direct relationship between the subdistribution hazard of the j-th event and the cumulative incidence function. The following lemma illustrates the relationship between the j-th cumulative incidence function and the j-th subdistribution hazard.

Lemma 14. *For the subdistribution hazard of the j-th event, we have*

$$\lambda_j^{SD}(t) = -\frac{d\log(1 - CIF_j(t))}{dt}.$$

Proof.

$$-\frac{d\log(1 - CIF_j(t))}{dt} = -\frac{d(1 - P(T \leq t; D = j)}{dt} =$$

$$= -\frac{d\log(P(T > t \bigcup D \neq j))}{dt} = \frac{dP(T \leq t; D = j)}{dt} \frac{1}{P(T > t \bigcup D \neq j)} =$$

$$\lim_{h \downarrow 0} \frac{P(t \leq T < t + h; D = j)}{hP(T > t \bigcup D \neq j)} = \lim_{h \downarrow 0} \frac{P(t \leq T < t + h; D = j | T > t \cup D \neq j)}{h}$$

□

Subdistribution and cause-specific hazard differ only in terms of which subjects are considered to be at risk. For subdistribution hazard, the risk process has a different form. In this section we follow ideas from Fine and Gray [1999]. The estimation of subdistribution hazard is divided into three possible data formats. Firstly, we will consider data without censoring. Secondly, we will generalise the concept to situations where we observe censoring even though the event has occurred. Finally, we will briefly introduce the most general case.

3.2.1 Complete Data

Assume now that there is no censoring in our dataset, as a consequence we directly observe vectors of independent observations

$$(T_1, D_1)^T, \ldots, (T_n, D_n)^T.$$

Let's introduce martingale notation used in upcoming theorems.

Definition 20 (Risk process for the j-th type of event)**.** *The* risk process of the i-th subject *for subdistribution hazard is defined as*

$$Y_{i,j}^{SD}(t) = 1 - N_{i,j}(t-) = \mathbf{1}\{T_i \geq t \cup D_i \neq j\}.$$

Martingale theorem

A martingale property is important for derivations of proportional hazard regression models. The next theorem will provide proof of this fact.

Theorem 15 (Martingale of subdistribution hazard)**.** *We have random vector $(T_i, D_i)^T$ with corresponding CS hazards $\lambda_1^{CS}(t), \ldots, \lambda_K^{CS}(t)$. Assume $P(D_i \neq j) > 0$. Then, it holds that*

$$M'_{i,j}(t) = N_{i,j}(t) - \int_0^t Y_{i,j}^{SD}(u)\lambda_j^{SD}(u)du$$

is a martingale with respect to the filtration

$$\sigma\left(N_{i,j}(u), Y_{i,j}^{SD}(u); u \leq t\right) = \mathcal{F}'^j.$$

Proof. The proof is similar to the proofs of Theorem 8 and Theorem 2, but some parts differ. To verify the martingale property, three properties have to be checked.
The martingale $M^{'}_{i,j}(t)$ consists of a counting process and a compensator. The sum of two measurable functions is a measurable function. The counting process is trivially measurable since it is in the generator of the σ-algebra. Regarding the compensator, heuristically we can argue that we have information up to time t. Based on this information, we are able to decide about value of the compensator. Integrability can be deduced because the counting process is bounded by one and from the following inequalities

$$\mathbb{E}|M^{'}_{i,j}(t)| \leq \mathbb{E}|N_{i,j}(t)| + \int_0^t \mathbb{E}\left|Y^{SD}_{i,j}(u)\lambda^{SD}_j(u)\right|du \leq$$

$$\leq 1 + \int_0^t P(T_i \geq u \cup D_i \neq j)\lambda^{SD}_j(u)du \leq$$

$$\leq 1 + \int_0^t \lim_{h \downarrow 0} P(T_i \geq u \cap D_i \neq j)\frac{P(u \leq T_i < u + h; D = j | T_i \geq u \cup D_i \neq j)}{h}du =$$

$$= 1 + \int_0^t \lim_{h \downarrow 0} \frac{P(u \leq T_i < u + h; D_i = j)}{h}du =$$

$$= 1 + \int_0^t \lim_{h \downarrow 0} \frac{P(u \leq T_i < u + h | D_i = j)P(D_i = j)}{h}du = 1 + P(D_i = j) \leq 2.$$

In this section, the definition of conditional expected value was gradually utilised in the final equation, and the property of density was further employed. Now, it remains to prove the martingale property.

So, let us have a set $F \in \mathcal{F}^{'j}_t$. We make disjoint decomposition of Ω and use intersection with the set F.

$$F = F \cap \Omega = F \cap ([T_i \leq t \cap D_i = j] \cup [T_i > t \cup D_i \neq j])$$

First, let's consider the first set (clearly $\mathcal{F}^{'j}_t$-measurable), which implies that event j occurred for subject i by time t. This means for the counting process that a jump has already occurred by time t, and it remains constant; hence, it holds for $s > 0$,

$$\int_{F \cap [T_i \leq t \cap D_i = j]} N_{i,j}(t + s) - N_{i,j}(t)dP = 0.$$

Similarly, we can deduce the integrals' nullity:

$$\int_{F \cap [T_i \leq t \cap D_i = j]} \int_t^{t+s} \mathbf{1}\{[T_i \geq u] \cup [D_i \neq j]\}\lambda^{SD}_j(u)dudP = 0.$$

This can be seen from the fact that the intersection of the set $[T_i \leq t \cap D_i = j]$ and the set that appears in the integrand's indicator is an empty set, yielding a trivial null integrand.

Now, let's take the second set from the decomposition, $[T_i > t \cup D_i \neq j]$ and again deduce that both integrals are equal.

The first important thing to note is that the set does not contain any proper (non-trivial) subset within $\mathcal{F}^{'j}_t$. For example, for $u > t$, consider the set

$$[T_i > u \cup D_i \neq j] \subset [T_i > t \cup D_i \neq j]\,.$$

However, this set does not belong to the filtration till time t. If the set was a subset, then $[T_i \leq u; D_i = j]$, would be in $\mathcal{F}_t^{'j}$. We can see that this set has to be either $\emptyset$ or Ω for $u > t$. Therefore, for $\omega \in [T_i > t \cup D_i \neq j]$, the following conditional expected value equals a constant:

$$\mathbb{E}\left[N_{i,j}(t+s) - N_{i,j}(t)|\mathcal{F}_t^{'}\right] = k.$$

Similarly, the following must hold:

$$\mathbb{E}\left[\int_t^{t+s} Y_{i,j}^{SD}(u)\lambda_j^{SD}(u)du|\mathcal{F}_t^{'}\right] = k^{'}.$$

If it's a martingale, both constants must be equal:

$$kP(T_i > t \cup D_i \neq j) = \int_{[T_i > t \cup D_i \neq j]} \mathbb{E}\left[N_{i,j}(t+s) - N_{i,j}(t)|\mathcal{F}_t^{'}\right] dP =$$

$$= \int_{[T_i > t \cup D_i \neq j]} N_{i,j}(t+s) - N_{i,j}(t) dP$$

$$= \int_\Omega \mathbf{1}\{[T_i > t \cup D_i \neq j]\}\left(N_{i,j}(t+s) - N_{i,j}(t)\right) dP.$$

Gradually, we utilised the definition of conditional expected value since the set over which we integrate is $\mathcal{F}_t^{'j}$ measurable. The final expression is merely the expected value of a binary random variable. Thus, we need to calculate the probability of the variable taking the value of one. It is clear in the indicator function. The difference between counting processes is non-zero only when event j occurs between times t and $t+s$. With this consideration, we obtain that the expected value is equal to the expression:

$$P([T_i > t; t \leq T_i \leq t+s; D_i = j] \cup [D_i \neq j; D_i = j; t \leq T_i \leq t+s]).$$

The second set is an empty set, hence trivially equal to zero. This follows from the assumption of competing risks, as it is impossible for a subject to experience multiple types of events. We obtained

$$k \cdot P(T_i > t \cup D_i \neq j) = P([T_i > t; t \leq T_i \leq t+s; D_i = j]).$$

Now, let's proceed with $k^{'}$ and calculate it in a similar form using analogous steps:

$$k^{'} \cdot P(T_i > t \cup D_i \neq j) = \int_{[T_i > t \cup D_i \neq j]} \int_t^{t+s} Y_{i,j}^{SD}(u)\lambda_j^{SD}(u)du\, dP =$$

$$= \int_t^{t+s} \mathbb{E}\mathbf{1}\{[T_i > t \cup D_i \neq j]\}\mathbf{1}\{T_i \geq u \cup D_i \neq j\}\lambda_j^{SD}(u)du =$$

$$= \int_t^{t+s} P(T_i \geq u \cup D_i \neq j)\lambda_j^{SD}(u)du.$$

Since $P(T_i > t \cup D_i \neq j) > 0$, we can divide by this expression. To prove the equality of k and k', we can cancel the term $P(T_i > t \cup D_i \neq j)$. Martingale property is then equivalent to the equation

$$\mathbf{1}\{[T_i > t \cup D_i \neq j]\}\frac{P(t < T_i \leq t+s; D_i = j)}{P(T_i > t \cup D_i \neq j)} =$$

$$\mathbf{1}\{[T_i > t \cup D_i \neq j]\}\frac{\int_t^{t+s} P(T_i \geq u \cup D_i \neq j)\lambda_j^{SD}(u)du}{P(T_i > t \cup D_i \neq j)} \quad a.s.$$

Let's continue with expression in nominator

$$P(t < T_i \leq t+s; D_i = j) = \int_t^{t+s} -\frac{\partial P(T_i \geq v; D_i = j)}{\partial v}\Big|_{v=u} du.$$

For further progress, we need $P(T_i \geq u \cup D_i \neq j) > 0$. This simply follows from the assumption that $P(D_i \neq j) > 0$. With this adjustment, we obtain the final desired equality. It's important to realise that the integrand, which involves partial derivatives along with $P(T_i \geq u \cup D_i \neq j)$ in the denominator, represents the subdistribution hazard.

$$-\int_t^{t+s} \frac{P(T_i \geq u \cup D_i \neq j)}{P(T_i \geq u \cup D_i \neq j)}\frac{\partial P(T_i \geq v; D_i = j)}{\partial v}\Big|_{v=u} du =$$

$$= \int_t^{t+s} P(T_i \geq u \cup D_i \neq j)\lambda_j^{SD}(u)du.$$

This concludes the proof. □

Since we make the estimators from the whole dataset we need to modify the filtration

$$\sigma\left(N_{i,j}(u), Y_{i,j}^{SD}(u); u \leq t; i = 1,\ldots,n\right) = \mathcal{F}_t^{'j}.$$

Proportional model

Now let's introduce a similar model as was shown in Section 3.1, but in this situation for subdistribution hazard. Let us have a regression parameter $\boldsymbol{\alpha_j}$ for the j-th type of event and an explanatory variable vector $\boldsymbol{Z}(t)$. Furthermore, assume true value of the parameter $\boldsymbol{\alpha_j^0}$ and a baseline hazard $\lambda_{0,j}^{SD}$ such that it holds

$$\lambda_{\boldsymbol{Z},j}^{SD}(t) = \lambda_{0,j}^{SD}(t)\exp\left(\boldsymbol{Z}^T(t)\boldsymbol{\alpha_j^0}\right).$$

This is assumed for all three situations: Complete data, Complete censored data and General censored data. In this a little bit changed setting Theorem 15 still could be used, but it is needed to add one more assumption about covariates. This is performed by adding $\boldsymbol{Z}(t)$ to filtration $\mathcal{F}_t^{'j}$. Now we are assuming filtration

$$\sigma\left(N_{i,j}(u), Y_{i,j}^{SD}(u), Y_{i,j}^{SD}(u)\boldsymbol{Z_i}(u); u \leq t; i = 1,\ldots,n\right).$$

Let's introduce likelihood for complete data without censoring from which we obtain estimates of coefficients $\boldsymbol{\alpha_j}$. Note that in this part we are presenting results mentioned in Fine and Gray [1999].

Definition 21. *The function*

$$\mathbf{L}'(\boldsymbol{\alpha_j}) = \prod_{i=1}^{n}\prod_{s>0}\left(\frac{Y_{i,j}^{SD}(s)\exp\left(\boldsymbol{\alpha_j}^T\boldsymbol{Z_i}(s)\right)}{\sum_{l=1}^{n} Y_{l,j}^{SD}(s)\exp\left(\boldsymbol{\alpha_j}^T\boldsymbol{Z_l}(s)\right)}\right)^{\Delta N_{i,j}(s)}$$

is called the partial likelihood of subdistribution hazard, *where* $\Delta N_{i,j}(t)$ *denotes the change in the counting process at time* t.

Maximising this function provides estimate $\hat{\boldsymbol{\alpha}}_{\boldsymbol{j}}'$ for the coefficients $\boldsymbol{\alpha_j}$. Again there is defined logarithmic likelihood and the corresponding score function. There is a small change of notation since there is used other at risk process.

Similarly, as for the CS hazard model, we introduce a little bit changed notation

$$S_n^{'(k)}(\boldsymbol{\alpha_j}, t) = \frac{1}{n}\sum_{i=1}^{n} Y_{i,j}^{SD}(t)\boldsymbol{Z_i}^{\bigotimes k}(t)\exp\boldsymbol{\alpha_j}^T\boldsymbol{Z_i}(t),$$

and

$$\overline{\boldsymbol{Z}}_n'(\boldsymbol{\alpha_j}, t) = \frac{S_n^{'(1)}(\boldsymbol{\alpha}_j, t)}{S_n^{'(0)}(\boldsymbol{\alpha}_j, t)}.$$

Definition 22 (Logarithmic partial likelihood for j-th SD hazard for complete data)**.** *The function*

$$\ell'(\boldsymbol{\alpha_j}) = \sum_{i=1}^{n}\int_0^{\infty}\left[\boldsymbol{\alpha_j}^T\boldsymbol{Z_i}(s) - log(nS_n^{'(0)}(\boldsymbol{\alpha_j}, s))\right]dN_{i,j}(s)$$

is called logarithmic partial likelihood for j-th subdistribution hazard.

By differentiating of log partial likelihood with respect to parameter $\boldsymbol{\alpha_j}$ we obtain the score function.

Definition 23 (Score function of $j-$th SD hazard)**.**

$$\boldsymbol{U}_{n,j}'(\boldsymbol{\alpha_j}) = \sum_{i=1}^{n}\int_0^{\infty}\left[\boldsymbol{Z_i}(t) - \overline{\boldsymbol{Z}}_n'(\boldsymbol{\alpha_j}, t)\right]dN_{i,j}(t)$$

From Definition 23 can be seen the importance of Theorem 15. The score function can be transformed into a sum of integrals with respect to martingale and the central limit theorem can be used. Value of covariates should be added to filtration $\mathcal{F}^{'j}$. Now we can claim that

$$M_{i,j}'(t) = N_{i,j} - \int_0^t Y_{i,j}^{SD}(u)\lambda_{0,j}^{SD}(u)\exp\left(\boldsymbol{Z}^T(u)\boldsymbol{\alpha_j^0}\right)du$$

is martingale. With this fact and with some more assumptions can be proved asymptotic properties of the estimator $\hat{\boldsymbol{\alpha}}_j'$

$$\sqrt{n}(\hat{\boldsymbol{\alpha}}_j' - \boldsymbol{\alpha}_j) \xrightarrow{D} N_p(0, (V')^{-1}),$$

where V' is limiting covariance matrix

$$V' = \int_0^{\infty}\left[\frac{\boldsymbol{s}^{'(2)}(\boldsymbol{\alpha}_j^0, u)}{s^{'(0)}(\boldsymbol{\alpha}_j^0, u)} - \overline{\boldsymbol{z}}^{SD}(\boldsymbol{\alpha}_j^0, u)^{\bigotimes 2}\right]s^{'(0)}(\boldsymbol{\alpha}_j^0, u)\lambda_{j,0}(u)du,$$

$$\overline{\boldsymbol{z}}'(\boldsymbol{\alpha}_j^0, u) = \frac{\boldsymbol{s}'^1(\boldsymbol{\alpha}_j^0, u)}{s'^0(\boldsymbol{\alpha}_j^0, u)},$$

$$s'^{(k)}(\boldsymbol{\alpha}_j, u) = \lim_{n\to\infty} S_n'^{(k)}.$$

Unknown quantity V could be estimated based on Fine and Gray [1999] by

$$\hat{V}' = \sum_{i=1}^{n} \left[\frac{\boldsymbol{S}_n'^{(2)}(\hat{\boldsymbol{\alpha}}_j', T_i)}{S_n'^{(0)}(\hat{\boldsymbol{\alpha}}_j', T_i)} - \overline{\boldsymbol{Z}}_n'(\hat{\boldsymbol{\alpha}}_j', T_i)^{\bigotimes 2} \right] \mathbf{1}\{D_i = j\}.$$

3.2.2 Censoring complete data

Sometimes, censoring in a dataset results from administrative follow-up. If we can obtain the censoring time even if the failure occurred earlier, we can transform the problem into a complete data setting. However, we need a new modified risk process and counting process. Moreover, we need the analogic property of independence of censoring time. As a reminder, let's assume we observe data

$$(X_1, C_1, \mathbf{1}\{T_1 < C_1\}D_1)^T, \dots, (X_n, C_n, \mathbf{1}\{T_n < C_n\}D_n)^T.$$

Martingale theorem

Let's state a theorem similar to Theorem 15. It is useful to realise that in the filtration, all of the terms are observable. We assume just an administrative follow-up.

Theorem 16 (Martingale of subdistribution hazard with complete censoring data)**.** *Assume that $P(D_i \neq j) > 0$. Then, it holds that*

$$M_{i,j}^*(t) = \int_0^t \mathbf{1}\{[C_i \geq u]\} dN_{i,j}(u) - \int_0^t \mathbf{1}\{[C_i \geq u]\} Y_{i,j}^{SD}(u)\lambda_j^{SD}(u)du$$

is a martingale with respect to the filtration

$$\sigma\left(\mathbf{1}\{[C_i \geq u]\}, \mathbf{1}\{[C_i \geq u]\}N_{i,j}(u), \mathbf{1}\{[C_i \geq u]\}Y_{i,j}^{SD}(u); u \leq t\right) = \mathcal{F}_t^{*j}$$

if and only if it holds

$$\lambda_j^{SD}(t) = \lim_{h\downarrow 0} \frac{P(t \leq T_i < t+h; D = j | C_i \geq t; T_i \geq t \bigcup C_i \geq t; D_i \neq j)}{h}$$

Proof. We present just an idea of the proof of the martingale property. Let us have a set $F \in \mathcal{F}_t^{*j}$. Now it is decomposed into the following sets

$$[T_i \leq t \cap D_i = j \cap C_i \leq t] \bigcup [T_i > t \cup D_i \neq j \cup C_i > t] = \Omega.$$

Similar to the proof of Theorem 15, we consider both sets to assess the martingale property. We assume times in the range of $0 < t < t + s$. When we examine the first set in the decomposition, we can see that the martingale property holds. This is because either an event or censoring has occurred before time t which means that the counting process remains constant over the time interval $(t, t+s)$. Therefore,

$$\int_{F\cap[T_i\leq t\cap D_i=j\cap C_i\leq t]} \int_t^{t+s} \mathbf{1}\{[C_i \geq u]\} dN_{i,j}(u) dP = 0.$$

Furthermore, it is also true that

$$\int_t^{t+s} \mathbf{1}\{[C_i \geq u]\} Y_{i,j}^{SD}(u) \lambda_j^{SD}(u) du dP = 0,$$

since combining $C_i \leq t$ and $C_i \leq u > t$ results in a zero indicator in the integrand. We continue with the same steps and notation as in the proof of Theorem 15. First, it is essential to realise there does not exist any nontrivial subset. Therefore, we obtain the following equations:

$$kP\left(T_i > t \cup D_i \neq j \cup C_i > t\right) = \int_{[T_i > t \cup D_i \neq j \cup C_i > t]} \int_t^{t+s} \mathbf{1}\{[C_i \geq u]\} dN_{i,j}(u) dP =$$

$$= \mathbb{E}\mathbf{1}\{[T_i > t \cup D_i \neq j \cup C_i > t]\}\mathbf{1}\{[C_i > t]\}\left(N_{i,j}(\min(t+s, C_i)) - N_{i,j}(t)\right)$$
$$= P(t < T_i \leq t+s; T_i \leq C_i; D_i = j).$$

By the same procedure applied to compensator part of martingale, we get the equations:

$$k^{'} P(T_i > t \cup D_i \neq j \cup C_i > t) =$$

$$= \int_t^{t+s} \mathbb{E}\mathbf{1}\{[T_i > t \cup D_i \neq j \cup C_i > t]\}\mathbf{1}\{[C_i > u]\}\mathbf{1}\{[T_i \geq t \cup D_i \neq j]\}\lambda_j^{SD}(u) du$$

$$= \int_t^{t+s} P([T \geq u; C_i \geq u] \cup [D_i \neq j; C_i \geq u])\lambda_j^{SD}(u) du$$

Now we compare constants k and $k^{'}$ and we obtain the last equation which holds almost surely if and only if there is independent censoring:

$$P(t < T_i \leq t+s; T_i \leq C_i; D_i = j) = -\int_t^{t+s} \frac{\partial P(T_i \geq v; D_i = j; C_i \geq u)}{\partial v}\Big|_{v=u} du =$$

$$= -\int_t^{t+s} \frac{\partial P(T_i \geq v; D_i = j)}{\partial v}\Big|_{v=u} du.$$

□

Again since we work with the whole dataset of n subjects, we assume filtration

$$\mathcal{F}_t^{*j} =$$
$$\sigma\left(\mathbf{1}\{[C_i \geq u]\}, \mathbf{1}\{[C_i \geq u]\}N_{i,j}(u), \mathbf{1}\{[C_i \geq u]\}Y_{i,j}^{SD}(u); u \leq t; i = 1, \dots, n\right).$$

Proportional model

Let us assume a proportional model (identical to that used for complete data) for the subdistribution hazard of the j-th type of event

$$\lambda_j^{SD}(t) = \lambda_{j,0}^{SD}(t) \exp\left(\boldsymbol{Z}^T(t)\boldsymbol{\alpha_j^0}\right),$$

where $\boldsymbol{\alpha_j^0}$ is the true value of the parameter, which we want to estimate.

In this case, we can use Theorem 16. Again it is needed to add a value of covariates $\boldsymbol{Z}(t)$ to the filtration $\mathcal{F}_t^{*j}$. Now we are assuming filtration

$$\sigma\Big(\mathbf{1}\{[C_i \geq u]\}, \mathbf{1}\{[C_i \geq u]\}N_{i,j}(u), \mathbf{1}\{[C_i \geq u]\}Y_{i,j}^{SD}(u), \mathbf{1}\{[C_i \geq u]\}Y_{i,j}^{SD}(u)\boldsymbol{Z_i}(u);$$
$$u \leq t; i = 1, \dots, n\Big).$$

Let's introduce the partial likelihood of censoring complete data from which we obtain estimates of coefficients $\boldsymbol{\alpha_j}$. Note that in this part we are presenting an analogical procedure which was performed for the complete data situation. In the paper Fine and Gray [1999] it was only mentioned that it can be done in the same way. We present it for completeness.

Definition 24. *The function*

$$\mathbf{L}^*(\boldsymbol{\alpha_j}) = \prod_{i=1}^{n} \prod_{s>0} \left(\frac{\mathbf{1}\{[C_i \geq u]\} Y_{i,j}^{SD}(s) \exp\left(\boldsymbol{\alpha_j}^T \boldsymbol{Z_i}(s)\right)}{\sum_{l=1}^{n} \mathbf{1}\{[C_l \geq u]\} Y_{l,j}^{SD}(s) \exp\left(\boldsymbol{\alpha_j}^T \boldsymbol{Z_l}(s)\right)} \right)^{\Delta N_{i,j}(s)}$$

is called the partial likelihood of subdistribution hazard, *where* $\Delta N_{i,j}(t)$ *denotes the change in the counting process at time* t.

Maximizing this function provides estimates $\hat{\boldsymbol{\alpha}}_j^*$ for the coefficients $\boldsymbol{\alpha}_j$. Again, the logarithmic likelihood and the corresponding score function is defined. There is a small change of notation since another at-risk process is used.

$$S_n^{*(k)}(\boldsymbol{\beta}_j, t) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{[C_i \geq t]\} Y_{i,j}^{SD}(t) \boldsymbol{Z_i}^{\bigotimes k}(t) \exp \boldsymbol{\alpha_j}^T \boldsymbol{Z_i}(t),$$

and

$$\overline{\boldsymbol{Z}}_n^*(\boldsymbol{\alpha_j}, t) = \frac{\boldsymbol{S}_n^{*(1)}(\boldsymbol{\alpha}_j, t)}{S_n^{*(0)}(\boldsymbol{\alpha}_j, t)}.$$

Definition 25 (Logarithmic partial likelihood for j-th SD hazard for censoring complete data)**.** *The function*

$$\ell^*(\boldsymbol{\alpha_j}) = \sum_{i=1}^{n} \int_0^\infty \left[\boldsymbol{\alpha_j}^T \boldsymbol{Z_i}(s) - log(nS_n^{*(0)}(\boldsymbol{\alpha_j}, s))\right] dN_{i,j}(s)$$

is called logarithmic partial likelihood for j-th SD hazard.

By differentiating of log partial likelihood with respect to parameter $\boldsymbol{\alpha_j}$ we obtain the score function.

Definition 26 (Score function of j-th SD hazard)**.**

$$\boldsymbol{U}_{n,j}^*(\boldsymbol{\alpha_j}) = \sum_{i=1}^{n} \int_0^\infty \left[\boldsymbol{Z_i}(t) - \overline{\boldsymbol{Z}}_n^*(\boldsymbol{\alpha_j}, t)\right] dN_{i,j}(t).$$

From Definition 26 can be seen the importance of Theorem 16. Score function could be transformed into a sum of integrals with respect to martingale and the central limit theorem could be used. Now we can claim that

$$M_{i,j}^*(t) =$$

$$\int_0^t \mathbf{1}\{[C_i \geq u]\} dN_{i,j}(u) - \int_0^t \mathbf{1}\{[C_i \geq u]\} Y_{i,j}^{SD}(u) \lambda_{0,j}^{SD}(u) \exp\left(\boldsymbol{Z}^T(u) \boldsymbol{\alpha_j^0}\right) du$$

is martingale. With this fact and some other assumptions could be proved asymptotic properties of estimator $\hat{\boldsymbol{\alpha}}_j^*$, especially asymptotic distribution:

$$\sqrt{n}(\hat{\boldsymbol{\alpha}}_j^* - \boldsymbol{\alpha}_j) \xrightarrow{D} N_p(0, (V^*)^{-1}),$$

where V^* is the limiting covariance matrix

$$V^* = \int_0^\infty \left[\frac{\boldsymbol{s}^{*(2)}(\boldsymbol{\alpha}_j^0, u)}{s^{*(0)}(\boldsymbol{\alpha}_j^0, u)} - \overline{\boldsymbol{z}}^*(\boldsymbol{\alpha}_j^0, u)^{\bigotimes 2}\right] s^{*(0)}(\boldsymbol{\alpha}_j^0, u)\lambda_{j,0}(u)du,$$

$$\overline{\boldsymbol{z}}^*(\boldsymbol{\alpha}_j^0, u) = \frac{\boldsymbol{s}^{*1}(\boldsymbol{\alpha}_j^0, u)}{s^{*0}(\boldsymbol{\alpha}_j^0, u)},$$

$$s^{*k}(\boldsymbol{\alpha}_j, u) = \lim_{n\to\infty} S_n^{*(k)}.$$

Unknown quantity V^* can be estimated by

$$\widehat{V}^* = \sum_{i=1}^n \left[\frac{\boldsymbol{S}_n^{*(2)}(\hat{\boldsymbol{\alpha}}_j^*, X_i)}{S_n^{*(0)}(\hat{\boldsymbol{\alpha}}_j^*, X_i)} - \overline{\boldsymbol{Z}}_n^*(\hat{\boldsymbol{\alpha}}_j^*, X_i)^{\bigotimes 2}\right] \mathbf{1}\{D_i = j\}\mathbf{1}\{C_i > T_i\}.$$

3.2.3 General censored data

In the most general case, we observe either time to event (of any type) or censoring time, so that previous methods cannot be applied. Fine and Gray [1999] suggested estimation of SD hazard based on inverse probability of censoring weighting (IPCW). Derivations are not provided in this thesis, for details see Fine and Gray [1999]. For completeness of text, there is presented a basic idea and score function.

Now we are not able to get for all individuals $(X_i, C_i, \mathbf{1}\{T_i < C_i\}D_i)$,because when we observe for example competing event, we do not know censoring time. Denote knowledge of the vital status of i-th subject at time t as

$$r_i(t) = \mathbf{1}\{C_i \geq \min(T_i, t)\}.$$

When $r_i(t) = 1$, then counting processes $N_{i,j}(t)$ and $Y_{i,j}^{SD}(t)$ are observable. If $r_i(t) = 1$ in general we are not able to observe the counting processes. Thus $r_i(t)N_{i,j}(t)$ and $r_i(t)Y_{i,j}^{SD}(t)$ are always observable. From this, we can construct the weights for the score function

$$w_i(t) = r_i(t)\frac{\hat{G}(t)}{\hat{G}(\min(X_i, t)},$$

where $\hat{G}(t)$ is Kaplan-Meier estimate of censoring event, same as in the Proposition 3. Now we define the score function for the general censored data

Definition 27 (Score function of j-th SD hazard for general censored data)**.**

$$\boldsymbol{U}_{n,j}^{**}(\boldsymbol{\alpha_j}) = \sum_{i=1}^n \int_0^\infty \left[\boldsymbol{Z_i}(t) - \frac{\sum_{l=1}^n w_l(t)Y_{l,j}^{SD}(t)\boldsymbol{Z_l}\exp\left(\boldsymbol{Z_l}^T\boldsymbol{\alpha_j}\right)}{\sum_{l=1}^n w_l(t)Y_{l,j}^{SD}(t)\exp\left(\boldsymbol{Z_l}^T\boldsymbol{\alpha_j}\right)}\right] w_i(t)dN_{i,j}(t)$$

By finding solution of an equation $\boldsymbol{U}_{n,j}^{**}(\boldsymbol{\alpha_j}) = 0$ we obtain estimator of $\boldsymbol{\alpha_j}$. With some additional assumptions and derivations, it can be proved that the estimator is consistent and asymptotically normally distributed. For more details see Fine and Gray [1999].

4. Simulations

In this chapter, the procedure for generating competing risk data is shown. Two simulations are performed, one for nonparametric methods and one for semiparametric estimation.

Firstly, we sum up which data we want to generate. Assume we have one event of interest and one competing event. We want to generate

$$(T_1, D_1)^T, \ldots, (T_n, D_n)^T.$$

Independent censoring times are generated afterwards. We focus now on initial step how to generate $T_1, \ldots, T_n$, from the distribution given by random variable T when we know its CS hazard functions $\lambda_1^{CS}, \lambda_2^{CS}$. Since we know CS hazard functions, we know cumulative hazard functions and as a result, we obtain survival function $\exp\left(-\Lambda_1^{CS}(t) - \Lambda_2^{CS}(t)\right)$. Ideas presented here are based on Bender et al. [2005]. It is well known that random variable $F(T)$ has uniform distribution U over $(0, 1)$. It is easy to prove that $1 - U$ has uniform distribution over $(0, 1)$ too. By this we obtain

$$U \sim 1 - F(T) \sim \exp\left(-\Lambda_1^{CS}(T) - \Lambda_2^{CS}(T)\right).$$

Assume that we can find the inverse function to calculate T directly, and then we find the data-generating mechanism. Sometimes, we are able to calculate it analytically. Sometimes, we have to use numerical methods. As we have obtained time for the event, we need to find out which event happened. This is done by CS hazards and their interpretation.

For $t > 0, h > 0$ we have expression

$$\begin{aligned} P\left(D = 1 | T \in [t, t+h), T \geq t\right) &= \frac{P(T \in [t, t+h), D = 1 | T \geq t)}{P(T \in [t, t+h) | T \geq t)} \\ &= \frac{\lambda_1^{CS}(t)}{\lambda_1^{CS}(t) + \lambda_2^{CS}(t)}. \end{aligned} \tag{4.1}$$

From this equation we can see one more interpretation of CS hazard which was presented by Beyersmann et al. [2009]. If an event for i-th subject happens at time t then the probability of the first event is given by Equation (4.1). To get the type of event, a binomial trial with the probability of the first event given by Equation (4.1) has to be performed. This concludes the first possibility of how to generate data. This interpretation gives us a reason, why there was a suggestion to estimate CS hazards for all types of events (to see magnitude relative to the other events). Theoretically, we can now add censoring times and calculate simulations. We call it the first type data-generating procedure.

One possible approach to the competing risks problem is via latent times variables. This other point of view was not presented in this thesis, since there are problems with identifiability, for more see Putter et al. [2007] and consequently Tsiatis [1975]. Beyersmann et al. [2009] claims that this is the reason why competing risk data should not be generated via latent variables. However, according to the following derivations, we do not see any problems for using this method.

Assume each type of event has its own continuous time to event $\tilde{T}_1, \ldots, \tilde{T}_K$. By taking the minimum of all latent variables we obtain random variable

$$T = min(\tilde{T}_1, \ldots, \tilde{T}_K),$$

(and D indicator which latent variable was minimum) which is the same random variable as was used in the thesis. We have to know how to generate these latent variables according to the chosen scenario. Assume that the latent variables are independent. Take the CS hazard of the first event and calculate

$$\lambda_1^{CS}(t) = \lim_{h \downarrow 0} \frac{P(t \leq T < t+h; D=1|T \geq t)}{h} =$$

$$\lim_{h \downarrow 0} \frac{P(t \leq \tilde{T}_1 < t+h|min(\tilde{T}_1, \ldots, \tilde{T}_K) \geq t)}{h} =$$

$$\lim_{h \downarrow 0} \frac{P(t \leq \tilde{T}_1 < t+h| \cap_{k=1}^K \tilde{T}_k \geq t)}{h} = \lim_{h \downarrow 0} \frac{P(t \leq \tilde{T}_1 < t+h|\tilde{T}_1 \geq t)}{h}.$$

In the last equation, we used the fact that the latent variables are assumed to be independent. As a result, we obtained the marginal hazard of random variable $\tilde{T}_1$. By Theorem 1 we know that (marginal) hazard completely specifies (marginal) survival function. As a consequence, we know how to select the distribution from which we want to generate data. We must generate k-th latent variable from the distribution specified by k-th CS hazard. The last thing to do is to add censoring times to complete data generating algorithm. This will be called the second-type data-generating procedure. This sums up how the data are generated in the upcoming simulations.

4.1 Nonparametric

In this section, we explore efficiency of estimator presented in Definition 15. According to Monte Carlo principles, the following scenarios were repeated 1000 times.

- censoring situations 10%, 30% and 50% censored,
- sample sizes 50, 100, 1000 and 5000 subjects,
- settings of CS hazard of an event of interest $\lambda_1^{CS} \in \{0.5, 1, 1.5\}$,
- settings of CS hazard of a competing event $\lambda_2^{CS} \in \{0.5, 1, 1.5\}$.

To specify the first data-generating procedure, it is possible to find the inverse of the survival function analytically. However in this case we do not have to follow the instruction exactly. One could realise that we can generate the random variable T simply from an exponential distribution with rate parameter $\lambda_1^{CS} + \lambda_2^{CS}$. The other data generating procedure we can see that both $\tilde{T}_1$ and $\tilde{T}_2$ have exponential distribution with rate parameters λ_1^{CS} and λ_2^{CS}. Exponential distribution was chosen for a censoring time distribution. Since all of the variables are exponential, the rate parameter for the censoring was calculated analytically. As a result, the rate parameter is in the form

$$(\lambda_1^{CS} + \lambda_2^{CS})\frac{p_c}{1-p_c},$$

where p_c is optional probability of censoring.

The results of the performed simulation are presented in Figure 4.1. All the estimates in the pictures are plotted to 95% quantile of time to either the first or second event. For 50% of censoring, there is presented 95% Monte Carlo confidence interval. According to the plot, we can see that the estimator of the cumulative incidence function seems to be consistent. The curves coincide. The only situation where some problems could be seen is with a combination of a small sample size and a high probability of censoring. The estimate varies from the true cumulative incidence function. However, it differs only in the part where we usually do not observe any events of interest. However, this problem seems to be fixed by increasing the sample size. We obtained nearly the same results as for all the other scenarios. Due to this fact, we present just one scenario. All of the figures are presented in the appendix to show the whole results. The presented figures are calculated using the first data-generating mechanism. The simulations were calculated using the second "latent variable" generating mechanism. We obtained nearly the same results, so there is no point in presenting all the figures twice. For each simulated dataset the estimator of cumulative incidence was performed by R Core Team [2023], package *cmprsk* [Gray, 2022].

4.2 Semiparametric

As was presented in Section 3.1, the Cox proportional model for CS hazards is one possibility how to analyse data. In order to explore the efficiency of a model,

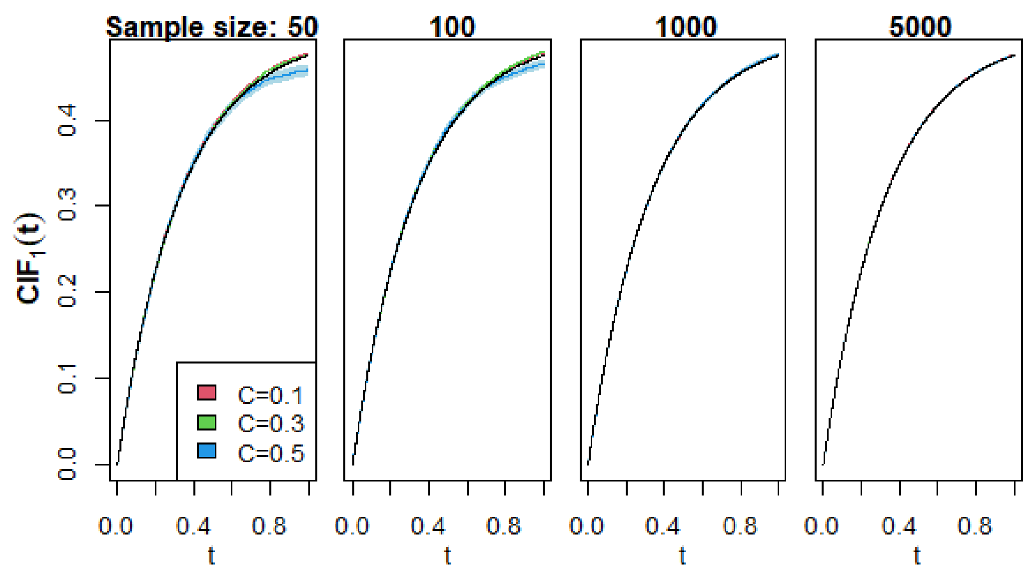

Figure 4.1: Nonparametric estimates of cumulative incidence function for scenario: $\lambda_1^{CS} = 1.5 \quad \lambda_2^{CS} = 1.5$.

Monte Carlo simulations of coefficients were performed. Assume that we have model given by Equation (3.1), where we assume two covariates (not time dependent) $\boldsymbol{Z} = (Z_1, Z_2)^T$. For Z_1 we assume a binary covariate with Bernoulli distribution with probability 1/2. The random variable Z_2 is a normal random variable independent of Z_1 with mean 50 and variance 12^2. It was simulated just for one event of interest and one competing event. There were made 1000 repetitions of each scenario. The simulation scenarios were the following:

- censoring situations 10%, 30% and 50% censored,
- sample sizes 50, 100 and 1000 subjects,
- regression parameters of an event of interest $\boldsymbol{\beta_1^0} = (0.3, -0.01)^T$,
- regression parameters of a competing event $\boldsymbol{\beta_2^0} = (0.1, 0.01)^T$.

To have a complete simulation study setting, it is necessary to specify the CS hazard of the event of interest and competing event for zero values of covariates. For this case, we assume just two scenarios, a constant scenario (both CS hazard functions are constant), a non-constant scenario inspired by Beyersmann et al. [2009]. We will explore each of these two scenarios more in detail. An exponential distribution with suitable rate parameters was chosen as the censoring distribution. The rate parameter was set up on smaller simulated data before the main simulation.

Constant initial cause-specific hazard

Assume that CS hazards of both types of events are constants for zero values of covariates(reference group). In the simulation, the initial CS hazard of the first event of the reference group is $\lambda_{0,1}^{CS}(t) = 0.5$, for the second type of an event this hazard is set to $\lambda_{0,2}^{CS}(t) = 0.2$. We follow the same notation stated in the Equation (3.1).

For the first type of data-generating procedure, the inverse of the survival function (dependent on covariates) is $S^{-1}(u) = -log(u)/(\lambda_{\boldsymbol{Z},1}^{CS} + \lambda_{\boldsymbol{Z},2}^{CS})$. Binomial trial is given by probability of first event $\lambda_{\boldsymbol{Z},1}^{CS}/(\lambda_{\boldsymbol{Z},1}^{CS} + \lambda_{\boldsymbol{Z},2}^{CS})$.

Generating via the latent variables approach is slightly easier. For given covariate $\boldsymbol{Z}$ we can use a random generator of exponential distribution for the first event with rate $\lambda_{\boldsymbol{Z},1}^{CS}$, and in the same way generate the latent variable for the second one.

Non-constant initial cause-specific hazard

For the first and second event, we set initial CS hazards as functions

$$\lambda_{0,1}^{CS}(t) = \frac{0.09}{t+1},$$

$$\lambda_{0,2}^{CS}(t) = 0.0002t.$$

As a consequence by simple integration and with the usage of the Lemma 4 we obtain the survival function

$$\exp\left(-\int_0^t \lambda^{CS}_{\boldsymbol{Z},1}(s)ds - \int_0^t \lambda^{CS}_{\boldsymbol{Z},2}(s)ds\right)$$

$$= \exp\left(-0.09\log(t+1)\exp\left(\boldsymbol{Z}^T\boldsymbol{\beta^0_1}\right) - 0.0002\frac{t^2}{2}\exp\left(\boldsymbol{Z}^T\boldsymbol{\beta^0_2}\right)\right).$$

We can see that the function is more complex and finding the inverse is not easy. We have to use a numerical method to calculate the inverse function.

The latent variables approach is again a little bit easier. The variable for the first event is generated from the distribution with the survival function

$$\tilde{S}_1(t) = \exp\left(-0.09\log(t+1)\exp(\boldsymbol{Z}^T\boldsymbol{\beta^0_1})\right).$$

It is easy to see that a function

$$\exp\left(\frac{-\log(u)}{0.09\exp(\boldsymbol{Z}^T\boldsymbol{\beta^0_1})}\right) - 1$$

is an inverse function to the survival function. As a consequence, we do not need to use numerical algorithms to solve the inverse function. Similarly, we could derive analogical results for the competing event. The estimators of regression coefficients were performed by fit of ordinary Cox model by R Core Team [2023], *survival* package Therneau [2023].

Results of the simulations

According to Figure 4.2 all of the confidence intervals are getting narrower and around the true value as the sample size increases, so all of the estimates seem to be consistent. For nonparametric simulations, we simulated all parameters using the first data-generating mechanism. Since the semiparametric simulation results were summarised in just one figure, we also present results for the second data-generating mechanism. We can see nearly the same results. There is a slight difference in competing events for both covariates for sample size 1000 subjects. For these situations, the confidence intervals are the same length for both data-generating procedures. For the latent variables, they are not that centred around the true value of the parameters. However, all of the confidence intervals cover the true value regardless of the type of data generation.

Beyersmann et al. [2009] suggested generating competing risk data by the first data-generating procedure rather than the second one. The authors argue that there is a problem with identifiability and that in practice it is not entirely appropriate to assume independence of latent variables. It is true that if we analyse the data and view cause-specific risk as the marginal risk of a given latent variable, this approach is often incorrect. However, given the simulations and derivations demonstrated above, if we have the opposite motivation to generate the data, then this procedure is applicable. The second data-generating procedure can often simplify the situation. For example the already mentioned computing the inverse function numerically or analytically.

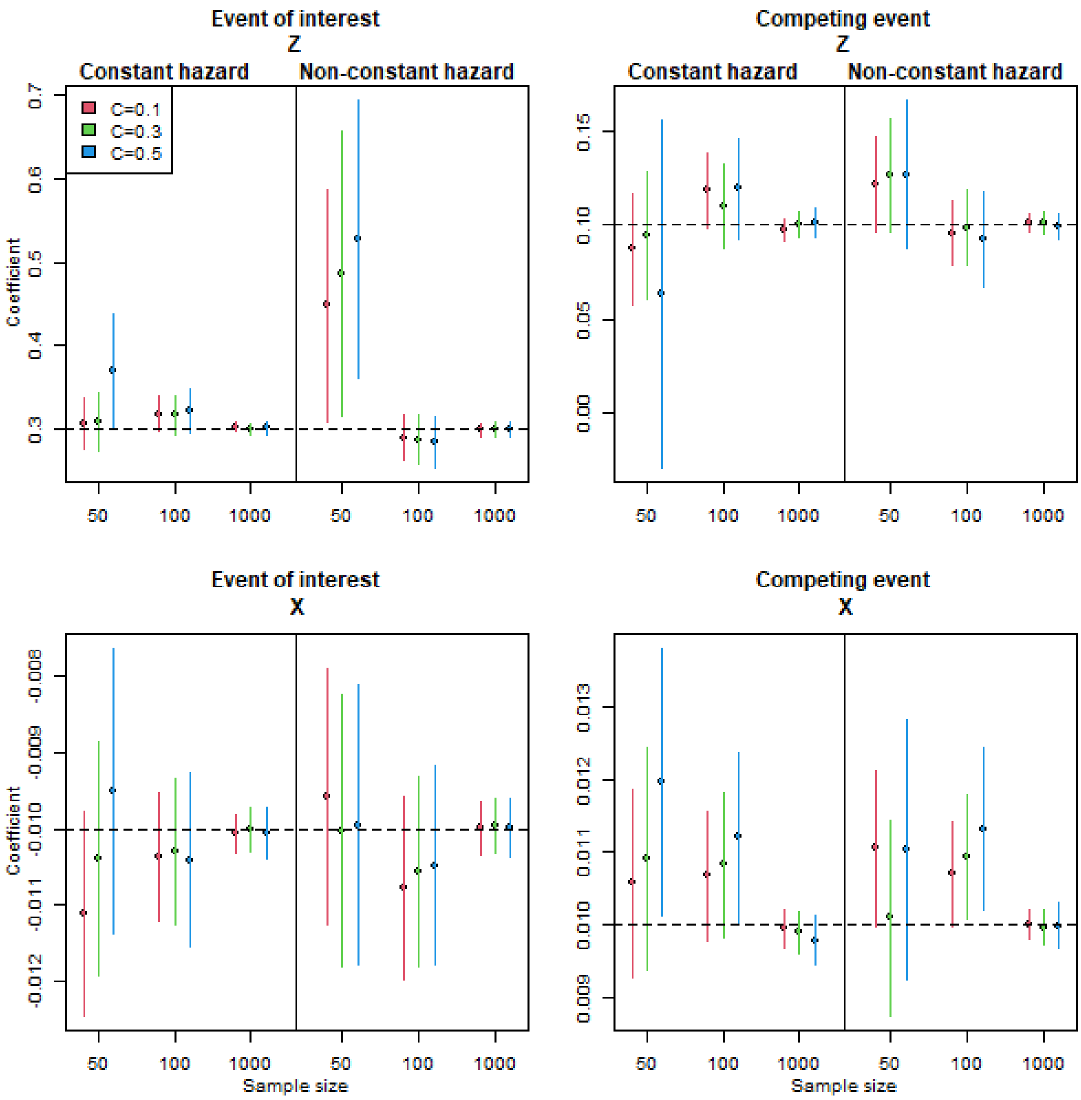

Figure 4.2: Monte Carlo 95% confidence intervals estimates of regression parameters

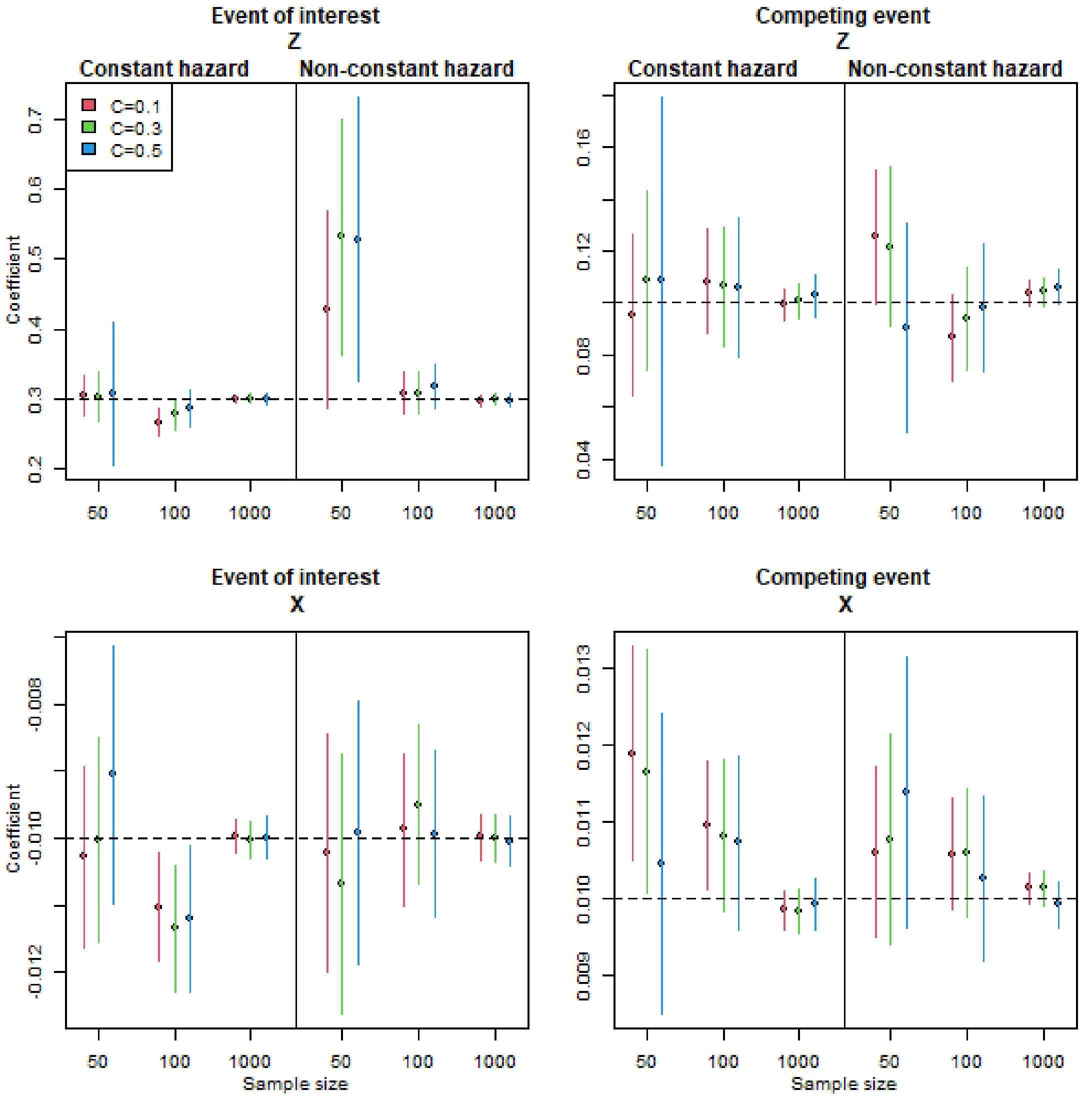

Figure 4.3: Monte Carlo 95% confidence intervals estimates of regression parameters by the second data generating mechanism

Conclusion

First, we presented the motivation for the problem of competing risks. Several practical examples were shown to illustrate where this phenomenon can occur.

In this thesis, we introduced elementary knowledge about competing risk models, both nonparametric and semiparametric. We defined basic concepts and derived elementary properties. Furthermore, we discussed an approach where we ignore competing risks and compared the possible bias of doing so.

The main result of the second chapter, and indeed the primary outcome of the entire master's thesis, is the proof of the asymptotic properties of the estimator of a cumulative incidence function (Theorem 13). This proof is performed with the necessary background in martingale theory and classical survival analysis, without employing the theory of multi-state models and product integrals.

In the third chapter, semiparametric methods were introduced. Firstly, we discussed the possibility of modelling cause-specific and subdistribution hazards. For the subdistribution hazard, we had to consider three possible situations: complete data, complete censoring data, and the most general case. Partial (log) likelihood and corresponding score functions were defined for both hazards and for all types of data. The main result of this chapter is the proof of Theorem 15 and the presentation of the idea of the proof of Theorem 16. The theorems proved a martingale property for distribution hazard. This property is crucial for further derivations of asymptotic properties.

A small simulation study in the last chapter demonstrated how well the cumulative incidence function can be estimated by the procedure presented in the second chapter. The results showed that with an increased number of subjects in the dataset, the estimator became more precise. Similarly, the Monte Carlo confidence intervals decreased with an increasing number of subjects.

In the second simulation study, we assumed that the Cox model holds for cause-specific hazards. We considered one binary and one continuous variable, thereby simulating two regression parameters. Once again, from the results, we were able to observe that with an increasing number of subjects, the estimates converged to the true values.

Bibliography

Odd Aalen. Nonparametric estimation of partial transition probabilities in multiple decrement models. *The Annals of Statistics*, 6(3):534–545, 1978.

Arthur Allignol, Martin Schumacher, Christoph Wanner, Christiane Drechsler, and Jan Beyersmann. Understanding competing risks: a simulation point of view. *BMC Medical Research Methodology*, 11:1–13, 2011.

Per K Andersen, Ornulf Borgan, Richard D Gill, and Niels Keiding. *Statistical Models Based on Counting Processes*. Springer Science & Business Media, 2012.

Peter C Austin, Douglas S Lee, and Jason P Fine. Introduction to the analysis of survival data in the presence of competing risks. *Circulation*, 133(6):601–609, 2016.

Ralf Bender, Thomas Augustin, and Maria Blettner. Generating survival times to simulate cox proportional hazards models. *Statistics in Medicine*, 24(11): 1713–1723, 2005.

Jan Beyersmann, Aurelien Latouche, Anika Buchholz, and Martin Schumacher. Simulating competing risks data in survival analysis. *Statistics in Medicine*, 28 (6):956–971, 2009.

Patrick Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, 2013.

Efron. The two sample problem with censored data. In *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics*, pages 831–853, Prentice-Hall, New York, 1967.

Jason P Fine and Robert J Gray. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446):496–509, 1999.

Thomas R Fleming and David P Harrington. *Counting Processes and Survival Analysis*. John Wiley & Sons, 2011.

Ted A Gooley, Wendy Leisenring, John Crowley, and Barry E Storer. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statistics in Medicine*, 18(6):695–706, 1999.

Bob Gray. *cmprsk: Subdistribution Analysis of Competing Risks*, 2022. URL `https://CRAN.R-project.org/package=cmprsk`. R package version 2.2-11.

John D Kalbfleisch and Ross L Prentice. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, 2011.

Bryan Lau, Stephen R Cole, and Stephen J Gange. Competing risk regression models for epidemiologic data. *American Journal of Epidemiology*, 170(2): 244–256, 2009.

DY Lin. Non-parametric inference for cumulative incidence functions in competing risks studies. *Statistics in Medicine*, 16(8):901–910, 1997.

Hein Putter, Marta Fiocco, and Ronald B Geskus. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*, 26(11):2389–2430, 2007.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023. URL `https://www.R-project.org/`.

Terry M Therneau. *A Package for Survival Analysis in R*, 2023. URL `https://CRAN.R-project.org/package=survival`. R package version 3.5-3.

Anastasios Tsiatis. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, 72(1):20–22, 1975.

List of Figures

A. Appendix

A.1 Weak Convergence

Some basic definitions and properties about the $D[0,\tau]$ space will be presented in this section. Space of stochastic processes with right continuous paths with left hand limits on $[0,\tau]$ will be denoted as $D[0,\tau]$.Consider a metric space $\mathcal{X}$ and the smallest sigma-algebra $\mathcal{B}$ that encompasses all the open sets within $\mathcal{X}$. A stochastic process, whose sample paths are within $\mathcal{X}$, can be defined as a measurable mapping from the $(\Omega,\mathcal{A}) \longrightarrow (\mathcal{X},\mathcal{B})$.

The metric that defines open sets on $D[0,\tau]$ is called Skorokhod metric. Let Φ be the set of all strictly increasing continuous functions f mapping $[0,\tau]$ onto $[0,\tau]$ so that$f(0)=0$ and $f(\tau)=\tau$

Definition 28 (Skorohod metrics)**.** *For any $g,h \in D[0,\tau]$ define*

$$d(g,h) = \inf\Big\{\varepsilon > 0 : \exists f \in \Phi \ s.t. \sup_{t\in[0,\tau]} |f(t)-t| \leq \varepsilon and \sup_{t\in[0,\tau]} |g(t)-h(f(t))| \leq \varepsilon\Big\}.$$

The distance d is called Skorokhod distance.

Definition 29. *Let P_n and P be probability measures on $(\mathcal{X},\mathcal{B})$. We say that P_n converges weakly to P as $n \longrightarrow \infty$, (denoted $Pn \overset{D[0,\tau]}{\longrightarrow} P$), if and only if $P_n(A) \longrightarrow P(A)$ for any $A \in \mathcal{B}$ such that $P(\partial A) = 0$, where ∂A is the boundary of the set A.*

When the sample space $\mathcal{X}$ is $\mathbb{R}^d$, weak convergence aligns with the convergence in distribution of a random vector X_n to a multivariate distribution P.

Theorem 17 (Continuous mapping theorem)**.** *Let h be a continuous mapping from a metric space $(\mathcal{X},\mathcal{B})$ to another metric space $(\mathcal{X}',\mathcal{B}')$, let $P_n \overset{D[0,\tau]}{\longrightarrow} P$ on $(\mathcal{X},\mathcal{B})$. Then*

$$P_n h^{-1} \overset{D[0,\tau]}{\longrightarrow} Ph^{-1}$$

on $(\mathcal{X}',\mathcal{B}')$.

For detailed derivations see Billingsley [2013].

A.2 Central limit theorems for sums of martingale integrals

In the thesis, the central limit theorem is used many times. The most important aspects are mentioned here. For more detailed derivations see Andersen et al. [2012, part II.5]

We will be working under the following conditions:

1. Let $\{N_{i,k}^{(n)} : k = 1,\dots,r, i = 1,\dots,n\}$ be a multivariate counting process with respect to the stochastic basis $(\Omega,\mathcal{A},\{\mathcal{F}_t\}t \geq 0, P)$.

2. Let the compensator $A_{i,k}^{(n)}$ for $N_{i,k}^{(n)}$ be continuous.

3. Let $H_{i,k}^{(n)}, k = 1, \ldots, r, i = 1, \ldots, n$, be bounded $\mathcal{F}_t-$predictable processes on the interval $[0, \tau]$.

Let $M_{i,k}^{(n)} = N_{i,k}^{(n)} - A_{i,k}^{(n)}$ be the $\mathcal{F}_t-$martingale for $N_{i,k}^{(n)}$.
Denote for $\varepsilon > 0$

$$U_{i,k;\varepsilon}^{(n)}(t) = \int_0^t H_{k,i}^{(n)}(u)\mathbf{1}\{H_{k,i}^{(n)}(u) > \varepsilon\}dM_{i,k}^{(n)}(u) \text{ and } U_{k;\varepsilon}^{(n)}(t) = \sum_{i=1}^{n} U_{i,k;\varepsilon}^{(n)}(t).$$

Denote

$$U_{i,k}^{(n)}(t) = \int_0^t H_{k,i}^{(n)}(u)dM_{i,k}^{(n)}(u) \text{ and } U_k^{(n)}(t) = \sum_{i=1}^{n} U_{i,k}^{(n)}(t).$$

All of these processes are square integrable martingales and it could be proved that the predictable variation of processes is equal to

$$\langle U_k^{(n)}; U_k^{(n)}\rangle(t) = \sum_{i=1}^{n} \int_0^t \left[H_{k,i}^{(n)}(u)\right]^2 dA_{i,k}^{(n)}(u),$$

and similarly

$$\langle U_{k;\varepsilon}^{(n)}; U_{k;\varepsilon}^{(n)}\rangle(t) = \sum_{i=1}^{n} \int_0^t \left[H_{k,i}^{(n)}(u)\right]^2 \mathbf{1}\{H_{k,i}^{(n)}(u) > \varepsilon\}dA_{i,k}^{(n)}(u).$$

Theorem 18 (Central limit theorem)**.** *Let for all $t \in [0, \tau]$ and all $k = 1, \ldots, r$*

$$\langle U_k^{(n)}; U_k^{(n)}\rangle(t) \xrightarrow{P} \int_0^t f_k^2(u)du < \infty$$

as $n \longrightarrow \infty$, where f_k are non-negative measurable functions and, for all $\varepsilon > 0$,

$$\langle U_{k;\varepsilon}^{(n)}; U_{k;\varepsilon}^{(n)}\rangle(t) \xrightarrow{P} 0.$$

Then

$$\left(U_1^{(n)}, \ldots, U_r^{(n)}\right) \xrightarrow{D^r[0,\tau]} \left(\int_0^t f_1 dW_1, \ldots, \int_0^t f_r dW_r\right),$$

where $W_1, \ldots, W_r$ are independent Brownian motions.

Proof. See Andersen et al. [2012, part II.5] □

A.3 Functional delta theorem

We want to explore convergence of random elements in some sample space. We will present a generalisation of differentiability.

Definition 30 (Hadamard differentiability). *Assume that we have* B *and* $B^{'}$ *Banach spaces then we call* $\phi : B \longrightarrow B^{'}$ *Hadamard differentiable at a point* $\theta \in B$ *if and only if a continuous linear map*

$$d\phi(\theta) : B \longrightarrow B^{'}$$

exists such that for all real sequences $a_n \longrightarrow \infty$ *and all convergent sequences* $h_n \longrightarrow h \in B$*,*

$$a_n \left(\phi(\theta + a_n^{-1} h_n) - \phi(\theta) \right) \longrightarrow d\phi(\theta)h \text{ as } n \longrightarrow \infty.$$

Theorem 19. *Let* T_n *be a sequence of random elements of* B*,* a_n *real sequence, such that*

$$a_n (T_n - \theta) \xrightarrow{w} Z$$

for some fixed point theta and random element Z *from* B*. (it is weak convergence of measures, in our context it is convergence in Skorochod metrics for* $D[0, \tau]$ *spaces) Suppose* $\phi : B \longrightarrow B^{'}$ *is Hadamard differentiable at* θ*. Then*

$$a_n (\phi(T_n) - \phi(\theta)) \xrightarrow{w} d\phi(\theta)Z$$

For more details see Andersen et al. [2012, part II.8]

A.4 Results of the simulations

Figures with results of the simulation from Section 4.1.

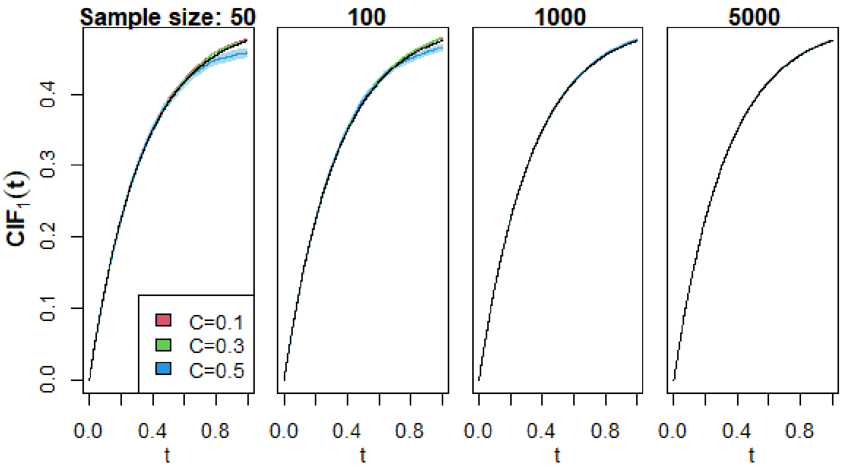

Figure A.1: Scenario: $\lambda_1^{CS} = 1.5 \quad \lambda_2^{CS} = 1.5$

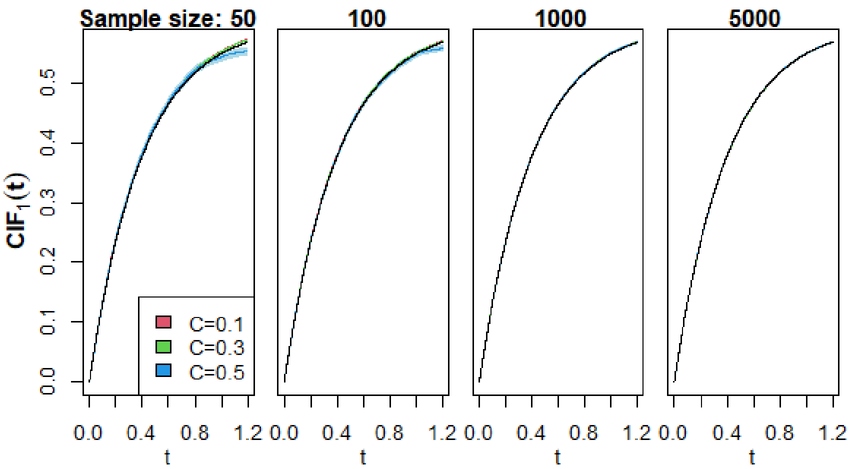

Figure A.2: Scenario: $\lambda_1^{CS} = 1.5 \quad \lambda_2^{CS} = 1$

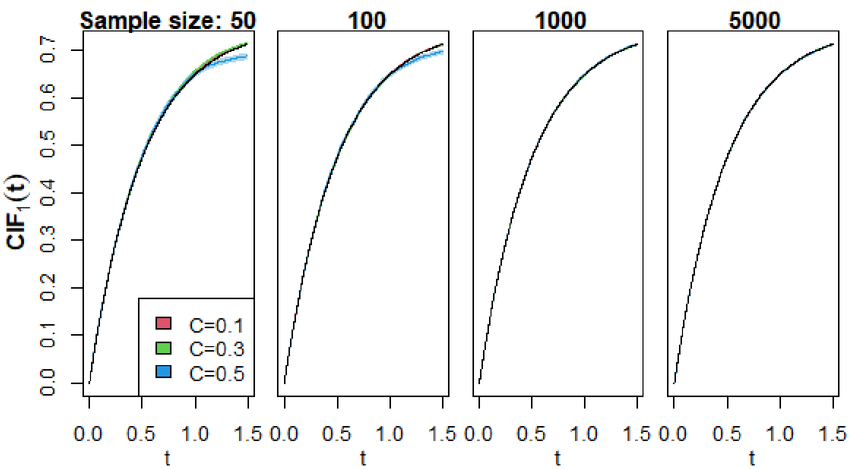

Figure A.3: Scenario: $\lambda_1^{CS} = 1.5 \quad \lambda_2^{CS} = 0.5$

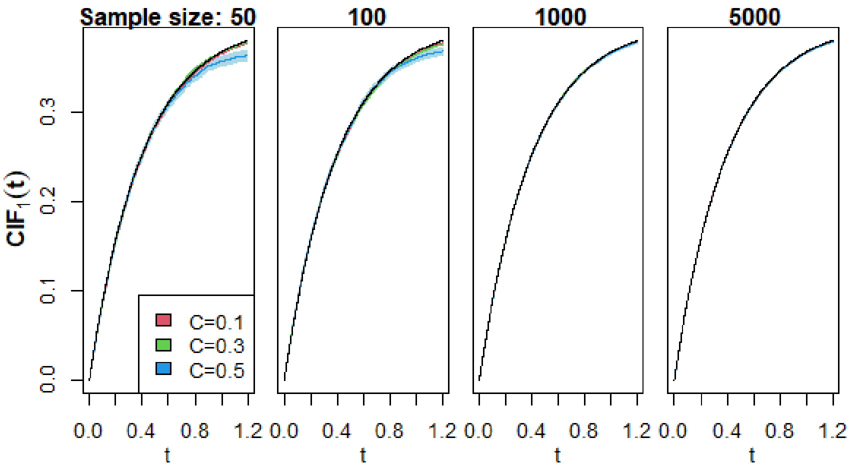

Figure A.4: Scenario: $\lambda_1^{CS} = 1 \quad \lambda_2^{CS} = 1.5$

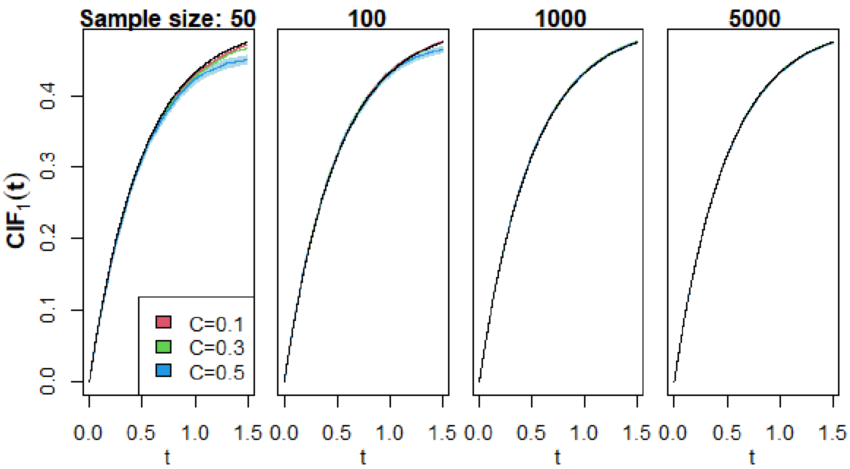

Figure A.5: Scenario: $\lambda_1^{CS} = 1 \quad \lambda_2^{CS} = 1$

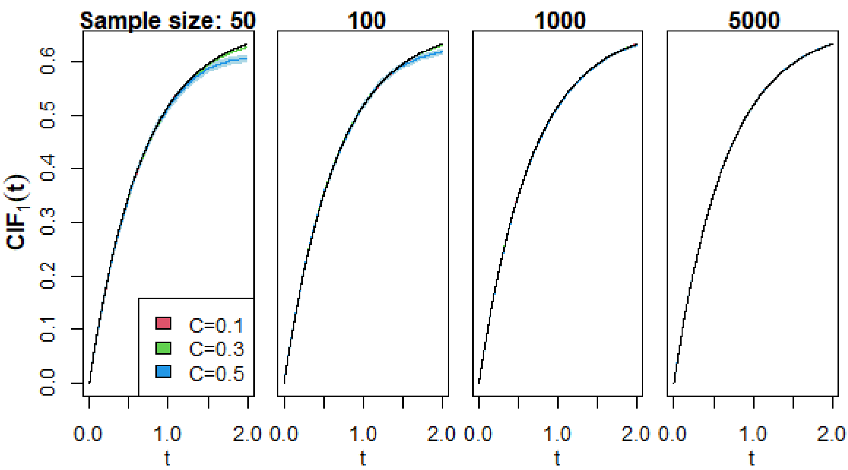

Figure A.6: Scenario: $\lambda_1^{CS} = 1 \quad \lambda_2^{CS} = 0.5$

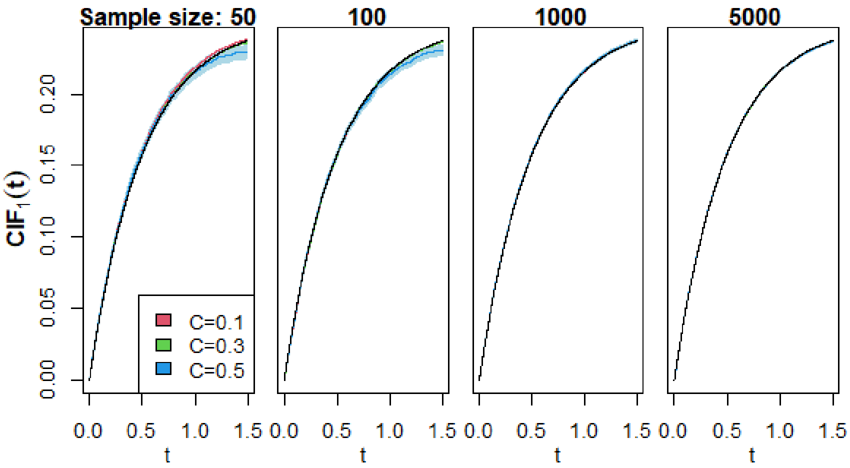

Figure A.7: Scenario: $\lambda_1^{CS} = 0.5 \quad \lambda_2^{CS} = 1.5$

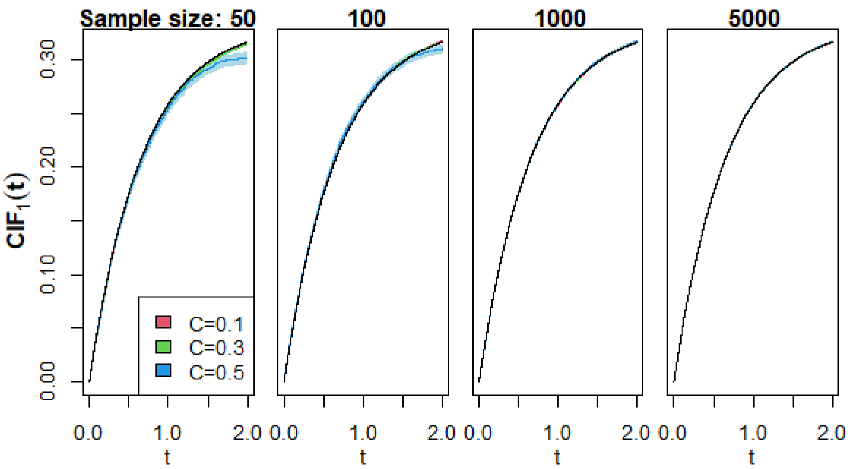

Figure A.8: Scenario: $\lambda_1^{CS} = 0.5 \quad \lambda_2^{CS} = 1$

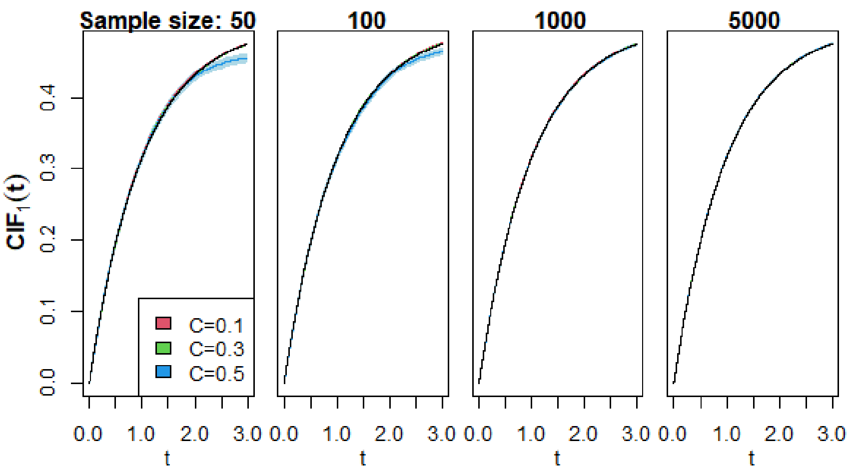

Figure A.9: Scenario: $\lambda_1^{CS} = 0.5 \quad \lambda_2^{CS} = 0.5$