



Semiparametric inference of competing risks data with additive hazards and missing cause of failure under MCAR or MAR assumptions

Laurent Bordes, Jean-Yves Dauxois, Pierre Joly

► To cite this version:

Laurent Bordes, Jean-Yves Dauxois, Pierre Joly. Semiparametric inference of competing risks data with additive hazards and missing cause of failure under MCAR or MAR assumptions. 2014. hal-00501865v2

HAL Id: hal-00501865

<https://hal.science/hal-00501865v2>

Preprint submitted on 3 Jul 2011 (v2), last revised 22 Jan 2014 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Semi-nonparametric inference of lifetime data under competing risks when failure cause is missing completely at random

L. Bordes^a, J.Y. Dauxois^{b,*}, P. Joly^{c,d}

^a*Laboratoire de Mathématiques Appliquées, UMR CNRS 5142, Université de Pau et des Pays de l'Adour, France*

^b*Laboratoire de Mathématiques, UMR CNRS 6623, Université de Franche-Comté, France*

^c*Université de Bordeaux, ISPED Bordeaux, France*

^d*INSERM U897, Epidémiologie et Biostatistique, Bordeaux, France*

Abstract

In this paper we consider a competing risk model under right censoring when the failure cause is missing completely at random. Two types of models are considered: a semiparametric one with additive hazard rate and a nonparametric one. In each case, preliminary estimators of the unknown parameters are obtained using mainly the lifetimes with known cause of failure. Then we show that the information given by the lifetimes with unknown failure cause can be optimally used to improve our estimates. The large sample behavior of our estimators is obtained and their performance on finite sample sizes illustrated through a simulation study.

Keywords: Additive hazards, competing risks, counting processes, missing failure cause, reliability, survival analysis.

1. Introduction

The competing risks models are useful in Survival Analysis or in Reliability in order to take into account the different causes of death of a patient, or the

*Corresponding author, Phone (+33)3 81 66 63 25, Fax (+33)3 81 66 66 23.

Email addresses: laurent.bordes@univ-pau.fr (L. Bordes),
jean-yves.dauxois@univ-fcomte.fr (J.Y. Dauxois), pierre.joly@isped.u-bordeaux2.fr
(P. Joly)

different causes of failure of a device. Based on possibly censored observation of the lifetimes and the indicators of cause of death, one can carry out a parametric, semi-parametric or nonparametric statistical inference. But, in some situations, it may happen that the cause of death or failure is missing for some individuals.

This paper deals with the analysis of lifetimes under competing risks when the indicator is missing completely at random. We consider an additive hazard rate model that allows to analyze such data from a semiparametric point of view in presence of covariates. The case without covariate is also considered from a nonparametric point of view. We need to emphasize that our model extends some existing models in the semi-nonparametric direction. However, the missingness mechanism is accounted here in the simplest way since it is supposed that the failure cause is missing completely at random (MCAR).

The problem of competing risks is not new. During the last two decades many models have been proposed in order to account for different causes of death or failure (see Crowder [7], for a large overview on the topic). In a number of real applications of competing risks models the authors have to face the problem of missing information: see e.g. Miyakawa [20], Usher and Hodgson [25], Lin *et al.* [16], Schabe [23], Goetghebuer and Ryan [14], Guttman *et al.* [15], Reiser *et al.* [22], Basu *et al.* [4], Flehinger *et al.* [10], Craiu and Duchesne [5], Craiu and Reiser [6].

Many authors have developed methods with accurate modelling of the missingness mechanism. Most of these works are based on parametric models. When a latent variable represents the missingness mechanism, an EM-type algorithm can be used to estimate the model parameters. In Craiu and Duchesne [5] such a procedure is proposed when the missingness mechanism depends both on the failure cause and the failure time. Recently, Craiu and Reiser [6] considered a very complete parametric model including dependence with the failure causes.

Some authors developed estimation procedures in the semi-nonparametric framework for two or more failure causes (see *e.g.* Miyakawa [20], Dinse [9], Lo [18], Schabe [23]). The special case of a possibly censored single failure cause differs from the competing risks model only by the fact that, in this case,

the censoring time is not an event of interest. However, when the censoring information is missing, we are close to the competing risks situation where failure causes are possibly missing. Some specific methods have been derived for various models including or not covariates, and several missingness mechanisms, see *e.g.* Gijbels *et al.* [12], McKeague and Subramanian [19], van der Laan and McKeague [26], Zhou and Sun [28] and Subramanian [24].

Goetghebeur and Ryan [14] proposed a competing risks model with proportional hazards assumption for the different failure causes. In their model, the mechanism of missingness may depend on the failure time (this is the missing at random assumption) while in our case it is independent of everything (this is the missing completely at random assumption). In our model each specific hazard rate function has its own semiparametric additive model and, contrary to Goetghebeur and Ryan [14], these functions are not linked.

The paper is organized as follows. In Section 2, the model is introduced and seen as a specific case of a nonhomogeneous Markov process. In Section 3, estimators of the Euclidean parameters are obtained using estimating equations. We also show that the observations with unknown failure cause are usable to estimate the sum of the previous parameters. Then, we develop a method that allows to account this information in an optimal way with respect to an asymptotic efficiency criterium. In Section 4, we show that our two types of parameter estimators are consistent and asymptotically gaussian. Section 5 deals with the estimation of the functional parameters. Here also we show that we can improve the asymptotic efficiency of our estimators using, in an optimal way, the information carried by the data with missing cause. These estimators are also proved to be consistent and asymptotically gaussian. Consistent estimators of the asymptotic variances are provided in each section. Finally, a Monte Carlo study is performed in Section 6 in order to assess the behavior of our euclidean parameter estimators on finite sample sizes.

2. Framework

Consider a population in which each individual is liable to die from any of $p \geq 2$ causes. The causes are not necessarily independent but we assume that each death is due to a single cause. Let us denote by T the individual lifetime and $d \in \{1, \dots, p\}$ its cause of death. Suppose that our interest focuses on the effect of a time-varying covariate vector $Z(\cdot)$ of dimension k . More precisely, let $\mathbb{Z}(t) = \{Z(u); u \leq t\}$ be the history up to time t of this covariate process and assume that an additive hazard model holds on the cause specific hazard rate function, that is:

$$\lambda_j(t|\mathbb{Z}) = \lim_{h \rightarrow 0^+} \frac{1}{h} P(T \in [t, t+h[, d = j | T \geq t, \mathbb{Z}(t)) = \lambda_{0j}(t) + \beta_j^T Z(t), \quad t \geq 0, \quad (1)$$

for $j = 1, \dots, p$ where $\lambda_{0j}(\cdot)$ is the baseline j th cause specific hazard rate function and $\beta_j \in \mathbb{R}^k$ is the vector of regression parameters associated to the j th cause.

Suppose also that the lifetime T is right-censored by a random variable (r.v.) C and write

$$X = T \wedge C \equiv \min(T, C) \text{ and } \delta = I(T \leq C),$$

where $I(\cdot)$ is the set indicator function. Let $\lambda_C(\cdot)$ denote the hazard rate function of the r.v. C and assume that conditionally on \mathbb{Z} , the r.v. C is independent from (T, d) .

Of course the futur cause of death is not known if C is observed instead of T . But in some situations it may happen that d is also not known even if T is observed. Let M denote the missingness indicator, i.e. $M = 1$ if the cause is known and $M = 0$ otherwise. Thus we are in a situation where the available observation for an individual is

$$(X, \delta, D, \mathbb{Z}(X)),$$

where $D = \delta M d$ reveals the failure cause d when the failure time is uncensored ($\delta = 1$) and $M = 1$ and is equal to zero otherwise. In the following we assume

Figure 1: Markov graph associated to (X, δ, D)

that the missing mechanism is such that:

$$P(M = 1|X, \mathbb{Z}, \delta = 1) = P(M = 1|\delta = 1) = \alpha \in [0, 1],$$

where α is an unknown parameter, and

$$P(M = 0|X, \mathbb{Z}, \delta = 0) = P(M = 0|\delta = 0) = 1.$$

One can see the observation of the vector (X, δ, D) , conditionally on \mathbb{Z} , as the realization of a $(p + 3)$ -states nonhomogeneous Markov process (see Fig. 1) with space set $\{0, 1, \dots, p, m, c\}$: state 0 is the initial state; state i , for $i = 1, \dots, p$, corresponds to the observation of the lifetime T with known cause of death i ; state m to the observation of the lifetime with missing cause; state c to a censored observation. Except 0, all the states are absorbant. From the assumptions on M and the independence between (T, d) and C , conditionally on Z , one can easily get that the instantaneous transition rates of this Markov process are, conditionally on \mathbb{Z} :

$$\begin{cases} \lambda'_j(t|\mathbb{Z}) &= \alpha \lambda_j(t|\mathbb{Z}), \text{ for } j \in \{1, \dots, p\}, \\ \lambda'_m(t|\mathbb{Z}) &= (1 - \alpha) \sum_{j=1}^p \lambda_j(t|\mathbb{Z}), \\ \lambda'_c(t|\mathbb{Z}) &= \lambda_C(t|\mathbb{Z}). \end{cases}$$

[Figure 1 here]

It is important to note that, up to a multiplicative constant (α or $1 - \alpha$), the additive form of the instantaneous rates is preserved (except for the transition $0 \rightarrow c$). It is obvious for $\lambda'_j(t|\mathbb{Z})$ with $1 \leq j \leq p$ but also for $\lambda'_m(\cdot|\mathbb{Z})$ since

$$\lambda'_m(t|\mathbb{Z}) = (1 - \alpha) (\lambda_{0m}(t) + \beta_m^T Z(t)),$$

where $\lambda_{0m}(\cdot) = \sum_{j=1}^p \lambda_{0j}(\cdot)$ et $\beta_m = \sum_{j=1}^p \beta_j$. This will help us to estimate the regression parameters β_1, \dots, β_p .

Now, let us suppose that we observe a sample $(X_i, \delta_i, D_i, \mathbb{Z}_i(X_i))_{1 \leq i \leq n}$ of $(X, \delta, D, \mathbb{Z}(X))$. Let $\tau < +\infty$ be the upper bound of the interval of study which means that individuals are only observed on the time interval $[0, \tau]$. Let $N_{ij}(\cdot)$, for $j \in \{1, \dots, p, m\}$, be the counting processes defined by:

$$\begin{aligned} N_{ij}(t) &= 1(X_i \leq t, D_i = j) \text{ for } j \neq m, \\ \text{and } N_{im}(t) &= 1(X_i \leq t, \delta_i = 1, D_i = 0). \end{aligned}$$

Finally write $Y_i(t) = 1(X_i \geq t)$ the individual risk process, for $i = 1, \dots, n$.

From Andersen *et al.* [1] or Fleming and Harrington [11], we know that, conditionally on \mathbb{Z} , the processes $M_{ij}(\cdot)$, for $j = 1, \dots, p$, and $M_{im}(\cdot)$ defined respectively by

$$M_{ij}(t) = N_{ij}(t) - \int_0^t Y_i(s) \lambda'_j(s|\mathbb{Z}) ds \text{ and } M_{im}(t) = N_{im}(t) - \int_0^t Y_i(s) \lambda'_m(s|\mathbb{Z}) ds,$$

for $t \geq 0$, are zero mean martingales with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$ defined by

$$\mathcal{F}_t = \sigma\{N_{ij}(s), N_{im}(s), Y_i(s); s \leq t; 1 \leq i \leq n, j \in \{1, \dots, p\}\}.$$

3. Statistical Inference on the Euclidean parameters

3.1. Estimators

The finite dimensional parameters of our model are: the probability α to know the cause of death, and the regression parameters β_1, \dots, β_p of each cause specific hazard rate functions. Recall that $\alpha = P(M = 1|\delta = 1)$. Thus, one can estimate empirically α by the proportion of lifetimes with known cause of death among the uncensored lifetimes, that is:

$$\hat{\alpha} = \frac{\sum_{i=1}^n 1(D_i > 0)}{\sum_{i=1}^n 1(\delta_i = 1)} = \frac{\sum_{j=1}^p N_{\cdot j}(\tau)}{N_{\cdot\cdot}(\tau)},$$

where

$$N_{\cdot j}(t) = \sum_{i=1}^n N_{ij}(t) \text{ and } N_{\cdot\cdot}(t) = \sum_{j=1}^p N_{\cdot j}(t) + \sum_{i=1}^n N_{im}(t), \text{ for all } t.$$

Extending an approach proposed by Lin et Ying [17] in case of a single cause of death, one can estimate β_j , for $j = 1, \dots, p$, by the solution $\hat{\beta}_j$ of the estimating equation $\mathcal{U}_j(\beta, \hat{\alpha}, \tau) = 0$ where

$$\mathcal{U}_j(\beta, \hat{\alpha}, \tau) = \sum_{i=1}^n \int_0^\tau [Z_i(s) - \bar{Z}(s)] [dN_{ij}(s) - \hat{\alpha} \beta^T Z_i(s) Y_i(s) ds],$$

and

$$\bar{Z}(s) = \frac{\sum_{i=1}^n Y_i(s) Z_i(s)}{\sum_{i=1}^n Y_i(s)}.$$

Now, since it has been seen that the cause specific hazard rate function $\lambda'_m(\cdot)$ associated to a missing cause has an additive form too, one can also estimate β_m by the solution $\hat{\beta}_m$ of the estimating equation $\mathcal{U}_m(\beta, \hat{\alpha}, \tau) = 0$ where

$$\mathcal{U}_m(\beta, \hat{\alpha}, \tau) = \sum_{i=1}^n \int_0^\tau [Z_i(s) - \bar{Z}(s)] [dN_{im}(s) - (1 - \hat{\alpha}) \beta^T Z_i(s) Y_i(s) ds].$$

Closed-form expressions of these estimators are available and given below (see Equation (10)).

At this stage, we are in a situation where each parameter β_j , for $j = 1, \dots, p$, has its own estimator $\hat{\beta}_j$. But we also have an estimator $\hat{\beta}_m$ of their sum $\beta_m = \beta_1 + \dots + \beta_p$. It is of course of interest to use it in order to improve the estimation of the first parameters β_j , for $j = 1, \dots, p$. In this order, we suggest to find the linear transformation of our estimator $(\hat{\beta}_1^T, \dots, \hat{\beta}_p^T, \hat{\beta}_m^T)^T$ which will lead to an estimator of $(\beta_1^T, \dots, \beta_p^T)^T$ with minimum asymptotic variance. More precisely, let \mathcal{H} be the family of all the block matrices

$$H = \begin{pmatrix} H_{11} & H_{12} & \cdots & H_{1p} & H_{1m} \\ H_{21} & H_{22} & \cdots & H_{2p} & H_{2m} \\ \vdots & \vdots & & \vdots & \vdots \\ H_{p1} & H_{p2} & \cdots & H_{pp} & H_{pm} \end{pmatrix},$$

where the H_{ij} , for $i = 1, \dots, p$ and $j = 1, \dots, p, m$, are $k \times k$ real valued matrices,

such that

$$H \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \\ \beta_m \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \quad (2)$$

for all vectors β_1, \dots, β_p in \mathbb{R}^k and $\beta_m = \beta_1 + \dots + \beta_p$. Write

$$\hat{q}(H) = \text{trace}(H \hat{\Sigma}_{\hat{\beta}, \infty} H^T),$$

where $\hat{\Sigma}_{\hat{\beta}, \infty}$ is an estimator of the asymptotic variance-covariance matrix of $(\hat{\beta}_1^T, \dots, \hat{\beta}_p^T, \hat{\beta}_m^T)^T$. With

$$\hat{H} = \underset{H \in \mathcal{H}}{\text{argmin}} \hat{q}(H),$$

an estimator of $(\beta_1^T, \dots, \beta_p^T)$ with minimal asymptotic variance is given by:

$$\begin{pmatrix} \tilde{\beta}_1 \\ \vdots \\ \tilde{\beta}_p \end{pmatrix} = \hat{H} \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \\ \hat{\beta}_m \end{pmatrix}.$$

Thus, the optimal estimators of the regression parameters are

$$\tilde{\beta}_i = \sum_{j=1}^p \hat{H}_{ij} \hat{\beta}_j + \hat{H}_{im} \hat{\beta}_m, \text{ for } i = 1, \dots, p,$$

and are called T -optimal estimators in the sequel. Such a way to improve estimators has been considered in Balakrishnan *et al.* [3].

Note that constraints on H , given by (2), are linear and do not link its lines. Indeed, denoting by I_k the identity matrix of order k , these constraints may be written

$$\begin{cases} H_{ii} + H_{im} &= I_k \\ H_{ij} + H_{im} &= 0, \end{cases}$$

for $1 \leq i \leq p$ and $j \in \{1, \dots, p\} \setminus \{i\}$.

On the other hand, we have

$$\hat{q}(H) = \sum_{i=1}^p \text{trace}(H_{i\bullet} \hat{\Sigma}_{\hat{\beta}, \infty} H_{i\bullet}^T) = \sum_{i=1}^p \hat{q}_i(H)$$

where $\hat{q}_i(H) = \text{trace}(H_{i\bullet} \hat{\Sigma}_{\hat{\beta}, \infty} H_{i\bullet}^T)$ and $H_{i\bullet}$ is the i th line block of H .

Thus, it is sufficient to solve separately the following problems (P_i) , for $i = 1, \dots, p$:

$$(P_i) \quad \begin{cases} \text{Find } H_{i1}, \dots, H_{im} \text{ which minimize } \text{trace}(H_{i\bullet} \hat{\Sigma}_{\hat{\beta}, \infty} H_{i\bullet}^T) \\ \text{such that: } H_{ii} + H_{im} = I_k, \\ H_{ij} + H_{im} = 0, \text{ for } j \neq i. \end{cases}$$

3.2. Example for $p = k = 2$

As an example, let us show how to solve problems (P_1) and (P_2) when $p = k = 2$. Since in this case (P_1) and (P_2) are identical, we only have to consider problem (P_1) . Let us introduce other temporary notations. For $j = 1, 2, m$, write

$$H^{(j)} = H_{1j} = \begin{pmatrix} h_{11}^{(j)} & h_{12}^{(j)} \\ h_{21}^{(j)} & h_{22}^{(j)} \end{pmatrix}.$$

Let

$$\hat{Q} = \begin{pmatrix} \hat{\Sigma}_{\hat{\beta}, \infty} & 0 \\ 0 & \hat{\Sigma}_{\hat{\beta}, \infty} \end{pmatrix}$$

and

$$L = (h_{11}^{(1)}, h_{12}^{(1)}, h_{11}^{(2)}, h_{12}^{(2)}, h_{11}^{(m)}, h_{12}^{(m)}, h_{21}^{(1)}, h_{22}^{(1)}, h_{21}^{(2)}, h_{22}^{(2)}, h_{21}^{(m)}, h_{22}^{(m)}) = (l_1, \dots, l_{12}).$$

One can rewrite the function $\hat{q}_1(H)$ to be minimized like:

$$\hat{q}_1(H) = \text{trace} \left(\begin{pmatrix} H^{(1)} & H^{(2)} & H^{(m)} \end{pmatrix} \hat{\Sigma}_{\hat{\beta}, \infty} \begin{pmatrix} H^{(1)} & H^{(2)} & H^{(m)} \end{pmatrix}^T \right) = L \hat{Q} L^T.$$

Now, since the constraints on H are

$$\begin{cases} H^{(1)} + H^{(m)} &= I_2 \\ H^{(2)} + H^{(m)} &= 0 \end{cases} \Leftrightarrow \begin{cases} h_{ij}^{(1)} + h_{ij}^{(m)} &= 1 \text{ for } 1 \leq i, j \leq 2 \\ h_{ij}^{(2)} + h_{ij}^{(m)} &= 0 \text{ for } 1 \leq i, j \leq 2 \end{cases},$$

they can be rewritten $CL = d$ where

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix} \quad \text{and} \quad d = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

Thus the Lagrange function for this optimization problem with linear constraints is given by

$$\ell(L, \lambda) = \frac{1}{2} L^T \hat{Q} L + (CL - d)^T \lambda$$

where λ is a Lagrange multiplier vector. The optimal parameters \hat{L} and $\hat{\lambda}$ necessarily satisfy the first-order conditions

$$\begin{cases} \frac{\partial \ell}{\partial L}(L, \lambda) &= 0 &= \hat{Q} L + C^T \lambda \\ CL &= d \end{cases},$$

leading to

$$\hat{\lambda} = -[C \hat{Q}^{-1} C^T]^{-1} d,$$

and

$$\hat{L} = \hat{Q}^{-1} C^T [C \hat{Q}^{-1} C^T]^{-1} d.$$

It has to be noted that \hat{Q} is invertible whenever $\hat{\Sigma}_{\beta, \infty}$ is.

4. Asymptotic behavior of the regression parameters estimators

4.1. Additional notations and assumptions, preliminary results

In order to simplify the notation in the following derivations, let us denote by $p+1$ the index corresponding to a missing cause (previously denoted by m).

Let

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \\ \beta_{p+1} \end{pmatrix} \text{ and } \hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \\ \hat{\beta}_{p+1} \end{pmatrix}.$$

If z is a column vector in \mathbb{R}^k , let us write

$$x^{\otimes l} = \begin{cases} 1 & \text{when } l = 0, \\ z & \text{when } l = 1, \\ zz^T & \text{when } l = 2. \end{cases}$$

Finally, let us denote by $S_l(\cdot)$, for $l = 0, 1, 2$, the processes defined by

$$S_l(s) = \frac{1}{n} \sum_{i=1}^n Y_i(s) Z_i^{\otimes l}(s),$$

for $0 \leq s \leq \tau$ and for b in \mathbb{R}^k the process $S_3(\cdot; b)$ defined by

$$S_3(s; b) = \frac{1}{n} \sum_{i=1}^n Y_i(s) Z_i^{\otimes 2}(s) b^T Z_i(s).$$

From now on we make the following assumptions.

- A1. The probability α to get a known cause of death is strictly positive.
- A2. The upper bound τ of the time interval of study is such that

$$0 < \int_0^\tau \lambda_{0j}(s) ds < +\infty, \text{ for } j = 1, \dots, p,$$

and the covariate processes $Z_i(\cdot)$ are $(\mathcal{F}_t)_{t \geq 0}$ -predictable and uniformly bounded with respect to $i \geq 1$.

- A3. For $0 \leq l \leq 2$, there exists functions $s_k(\cdot)$ defined on $[0, \tau]$ such that

$$\max_{0 \leq l \leq 2} \sup_{s \in [0, \tau]} \|S_l(s) - s_l(s)\| \xrightarrow{P} 0, \text{ when } n \rightarrow +\infty.$$

Moreover, the function $s_0(\cdot)$ is bounded below by a positive real number.

A4. With the notations $a(u) = s_2(u) - s_1^{\otimes 2}(u)/s_0(u)$,

$$A(t) = \int_0^t a(u) du \text{ and } \theta(t) = \int_0^t [s_0(u)\lambda_{0p+1}(u) + \beta_{p+1}^T s_1(u)] du, \quad (3)$$

the matrix $A(\tau)$ is positive definite and the real number $\theta(\tau)$ is strictly positive.

The matrix

$$\int_0^\tau [S_2(s) - S_1^{\otimes 2}(s)/S_0(s)] ds$$

is also positive definite. Note that, from previous assumptions, it was already true asymptotically.

A5. For all $b \in \mathbb{R}^k$, let $S_4(\cdot; b)$ be the process defined by

$$S_4(s; b) = \frac{1}{n} \sum_{i=1}^n (b^T Z_i(s)) Y_i(s) (s_1(s) Z_i^T(s) / s_0(s)).$$

There exist functions $s_3(\cdot; b)$ and $s_4(\cdot; b)$ such that, for all $b \in \mathbb{R}^k$,

$$\max_{3 \leq l \leq 4} \sup_{s \in [0, \tau]} \|S_l(s; b) - s_l(s; b)\| \xrightarrow{P} 0, \text{ when } n \rightarrow +\infty.$$

A6. The following functions are integrable on $[0, \tau]$:

$$a(\cdot), a(\cdot)\lambda_{0j}(\cdot), s_0(\cdot)\lambda_{0j}(\cdot), s_1(\cdot), \frac{s_1^{\otimes 2}(\cdot)}{s_0^2(\cdot)}, s_3(\cdot; \beta_j) \text{ and } s_4(\cdot; \beta_j), \text{ for } j = 1, \dots, p.$$

Now, let us introduce, for all $b \in \mathbb{R}^k$, the processes $\mathcal{U}_j(b, \alpha_j, \cdot)$ defined, for $j = 1, \dots, p+1$ and $t \in [0, \tau]$, by:

$$\mathcal{U}_j(b, \alpha_j, t) = \sum_{i=1}^n \int_0^t (Z_i(s) - \bar{Z}(s)) (dN_{ij}(s) - \alpha_j b^T Z_i(s) Y_i(s) ds),$$

where

$$\alpha_j = \begin{cases} \alpha & \text{if } 1 \leq j \leq p, \\ 1 - \alpha & \text{if } j = p+1. \end{cases}$$

Note that values of these processes at $s = \tau$ appear in the estimating equations of Section 3.1.

It is easily seen that these processes can be rewritten like

$$\mathcal{U}_j(b, \alpha_j, \cdot) = \sum_{i=1}^n \int_0^\cdot (Z_i(s) - \bar{Z}(s)) dM_{ij}(s),$$

for $j = 1, \dots, p+1$. Hence they are local square integrable $(\mathcal{F}_t)_{t \geq 0}$ -martingales as sum of stochastic integrals of predictable and bounded processes with respect to local square integrable martingales.

Now let us introduce two technical results useful in the following section. Let us note that all the functional convergence results of this paper are considered in the Skorohod space of cadlag functions $D[0, \tau]$.

Proposition 4.1. *Under Assumptions A1-A6, the multivariate process $\mathbb{U}_n(\cdot)$ defined, for all $t \in [0, \tau]$, by*

$$\mathbb{U}_n(t) = \frac{1}{\sqrt{n}} \begin{pmatrix} \mathcal{U}_1(\beta_1, \alpha_1, t) \\ \vdots \\ \mathcal{U}_{p+1}(\beta_{p+1}, \alpha_{p+1}, t) \\ \sum_{j=1}^{p+1} \frac{\alpha_j^*}{\theta(\tau)} M_{\cdot j}(t) \end{pmatrix},$$

where $\alpha_j^* = 1 - \alpha$ for $1 \leq j \leq p$ and $\alpha_{p+1}^* = -\alpha$, converges weakly in $D^{k(p+1)+1}[0, \tau]$, as n tends to infinity, to a zero mean multivariate gaussian martingale $\mathbb{U}_\infty(\cdot)$ with covariance matrix defined for all $t \in [0, \tau]$ by

$$\Sigma_{\mathbb{U}_\infty}(t) = \begin{pmatrix} \Theta_1(t) & 0 & 0 & \xi_1(t) \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & \Theta_{p+1}(t) & \xi_{p+1}(t) \\ \xi_1^T(t) & \cdots & \xi_{p+1}^T(t) & \xi_\alpha(t) \end{pmatrix},$$

where, for $j = 1, \dots, p+1$,

$$\begin{aligned} \Theta_j(t) &= \alpha_j \int_0^t [a(s)\lambda_{0j}(s) + s_3(s; \beta_j) - s_4(s; \beta_j) - s_4^T(s; \beta_j) + (\beta_j^T s_1(s))s_1^{\otimes 2}(s)/s_0^2(s)] ds, \\ \xi_j(t) &= \kappa_j \frac{\alpha(1-\alpha)}{\theta(\tau)} A(t)\beta_j, \text{ with } \kappa_j = 1, \text{ for } 1 \leq j \leq p \text{ and } \kappa_{p+1} = -1, \\ \xi_\alpha(t) &= \frac{\alpha(1-\alpha)}{\theta^2(\tau)} \theta(t) \end{aligned}$$

and $A(\cdot)$ as well as $\theta(\cdot)$ are defined in (3).

Proof. As in Andersen and Gill [2], the main idea is to apply Rebolledo's Theorem (see Rebolledo [21], or Andersen *et al.* [1], p. 83–84). Since this theorem is now of classical use, we only derive here the limit of the predictable variation process associated to $\mathbb{U}_n(\cdot)$. This will give us the asymptotic variance-covariance matrix function $\Sigma_{\mathbb{U}_\infty}(\cdot)$ of $\mathbb{U}_n(\cdot)$.

On one hand, straightforward calculations show that we have, for $j = 1, \dots, p+1$,

$$\begin{aligned} \left\langle \frac{1}{\sqrt{n}} \mathcal{U}_j(\beta_j, \alpha, \cdot) \right\rangle(t) &= \frac{1}{n} \sum_{i=1}^n \int_0^t (Z_i(s) - \bar{Z}(s)) (Z_i(s) - \bar{Z}(s))^T Y_i(s) \lambda_j'(s|Z) ds \\ &= \alpha_j \int_0^t [S_2(s) - S_1^{\otimes 2}(s)/S_0(s)] \lambda_{0j}(s) ds \\ &+ \alpha_j \int_0^t [S_3(s; \beta_j) - S_4(s; \beta_j) - S_4^T(s; \beta_j) + (\beta_j^T S_1(s)) S_1^{\otimes 2}(s)/S_0^2(s)] ds + o_P(1), \end{aligned}$$

which, by Assumptions A2-A6, converges in probability, as n tends to infinity, to $\Theta_j(t)$ given in the Theorem.

On the other hand, since the martingales M_{ij} and $M_{i'j'}$ are orthogonal for all $1 \leq i, i' \leq n$ and $1 \leq j \neq j' \leq p+1$, we have:

$$\langle \mathcal{U}_j(\beta_j, \alpha_j, \cdot), \mathcal{U}_{j'}(\beta_{j'}, \alpha_{j'}, \cdot) \rangle(t) = 0,$$

whenever j is different from j' and for all $t \in [0, \tau]$. This justifies the null terms in the asymptotic covariance matrix $\Sigma_{\mathbb{U}_\infty}(t)$.

Now, thanks again to the orthogonality between martingales with different indices, it is easy to show that, for $1 \leq l \leq p+1$, we have

$$\frac{1}{n} \left\langle \mathcal{U}_l(\cdot), \sum_{j=1}^{p+1} \frac{\alpha_j^* M_{\cdot j}(\cdot)}{\theta(\tau)} \right\rangle(t) = \frac{\alpha_l^*}{n\theta(\tau)} \sum_{i=1}^n \int_0^t (Z_i(s) - \bar{Z}(s)) Y_i(s) \lambda_j'(s|Z) ds$$

which converges in probability to

$$\xi_l(t) = \kappa_l \frac{\alpha(1-\alpha)}{\theta(\tau)} A(t) \beta_l,$$

when n tends to infinity. Finally, with the same kind of arguments we have

$$\left\langle \frac{1}{\sqrt{n}} \sum_{j=1}^{p+1} \frac{\alpha_j^* M_{\cdot j}(\cdot)}{\theta(\tau)} \right\rangle(t) = \frac{1}{n} \sum_{j=1}^{p+1} \frac{\alpha_j^*}{\theta^2(\tau)} \sum_{i=1}^n \int_0^t Y_i(s) \lambda_j'(s|\mathbb{Z}) ds \xrightarrow{P} \xi_\alpha(t) = \frac{\alpha(1-\alpha)}{\theta(\tau)} \theta(t),$$

when n tends to infinity. \square

4.2. Large sample behaviour of $\hat{\alpha}$

Lemma 4.2. *Under Assumptions A1-A4, we have*

$$\sqrt{n}(\hat{\alpha} - \alpha) = \sum_{j=1}^{p+1} \frac{\alpha_j^*}{\theta(\tau)} \frac{1}{\sqrt{n}} M_{\cdot j}(\tau) + o_P(1). \quad (4)$$

Moreover $\sqrt{n}(\hat{\alpha} - \alpha)$ converges weakly to a $N(0, \alpha(1-\alpha)/\theta(\tau))$ distribution when $n \rightarrow +\infty$.

Remark. The second result of this lemma is straightforward and doesn't require the first step. It arises from an easy application of the central limit theorem. However the first result will be useful in the next section.

Proof. From the definition of $\hat{\alpha}$ we can write:

$$\begin{aligned} \hat{\alpha} &= \frac{\sum_{i=1}^n \sum_{j=1}^p \int_0^\tau dN_{ij}(s)}{\sum_{i=1}^n \sum_{j=1}^{p+1} \int_0^\tau dN_{ij}(s)} \\ &= \frac{\sum_{i=1}^n \sum_{j=1}^p \int_0^\tau dM_{ij}(s) + \sum_{i=1}^n \sum_{j=1}^p \int_0^\tau Y_i(s) \lambda_j'(s|\mathbb{Z}) ds}{\sum_{i=1}^n \sum_{j=1}^{p+1} \int_0^\tau dM_{ij}(s) + \sum_{i=1}^n \sum_{j=1}^{p+1} \int_0^\tau Y_i(s) \lambda_j'(s|\mathbb{Z}) ds} \\ &= \frac{\sum_{j=1}^p M_{\cdot j}(\tau) + \alpha \sum_{i=1}^n \sum_{j=1}^p \int_0^\tau Y_i(s) \lambda_j(s|\mathbb{Z}) ds}{M_{\cdot \cdot}(\tau) + \sum_{i=1}^n \sum_{j=1}^p \int_0^\tau Y_i(s) \lambda_j(s|\mathbb{Z}) ds} \\ &= \frac{\sum_{j=1}^p M_{\cdot j}(\tau) + \alpha n \int_0^\tau [S_0(u) \lambda_{0p+1}(u) + \beta_{p+1}^T S_1(u)] du}{M_{\cdot \cdot}(\tau) + n \int_0^\tau [S_0(u) \lambda_{0p+1}(u) + \beta_{p+1}^T S_1(u)] du}, \end{aligned}$$

where $M_{\cdot}(\cdot)$ denotes the process $\sum_{i=1}^n \sum_{j=1}^{p+1} M_{ij}(\cdot)$. It follows that

$$\begin{aligned}
& \sqrt{n}(\hat{\alpha} - \alpha) \\
&= \frac{\frac{1-\alpha}{\sqrt{n}} \sum_{j=1}^p M_{\cdot j}(\tau) - \frac{\alpha}{\sqrt{n}} M_{\cdot p+1}(\tau)}{\frac{1}{n} M_{\cdot}(\tau) + \int_0^\tau [S_0(u) \lambda_{0p+1}(u) + \beta_{p+1}^T S_1(u)] du} 1(N_{\cdot}(\tau) > 0) - \sqrt{n} \alpha 1(N_{\cdot}(\tau) = 0) \\
&= \frac{\frac{1-\alpha}{\sqrt{n}} \sum_{j=1}^p M_{\cdot j}(\tau) - \frac{\alpha}{\sqrt{n}} M_{\cdot p+1}(\tau)}{\frac{1}{n} M_{\cdot}(\tau) + \int_0^\tau [S_0(u) \lambda_{0p+1}(u) + \beta_{p+1}^T S_1(u)] du} + o_P(1).
\end{aligned}$$

Moreover, using Lengart's inequality (see *e.g.* [1]) and Assumption A3, it is easily seen that

$$\frac{1}{n} M_{\cdot}(\tau) \xrightarrow{P} 0, \text{ as } n \rightarrow +\infty. \quad (5)$$

These two last results with Assumptions A3 and A4 complete the proof of (4).

As said in the previous remark, the asymptotic normality of $\hat{\alpha}$ is straightforward. \square

4.3. Large sample behaviour of $\hat{\beta}$

Theorem 4.3. *Under Assumptions A1-A6, the random vector $\sqrt{n}(\hat{\beta} - \beta)$ is asymptotically gaussian, with zero mean and positive definite covariance matrix*

$$\Sigma_{\hat{\beta}, \infty} = \Sigma_1(\tau) \Sigma_{\mathbb{U}_\infty}(\tau) \Sigma_1^T(\tau),$$

where

$$\Sigma_1(\tau) = \begin{pmatrix} \frac{A^{-1}(\tau)}{\alpha} & 0 & \cdots & 0 & -\frac{\beta_1}{\alpha} \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \frac{A^{-1}(\tau)}{\alpha} & 0 & -\frac{\beta_p}{\alpha} \\ 0 & \cdots & 0 & \frac{A^{-1}(\tau)}{1-\alpha} & \frac{\beta_{p+1}}{1-\alpha} \end{pmatrix}.$$

Proof. Write

$$\hat{\alpha}_j = \begin{cases} \hat{\alpha} & \text{if } 1 \leq j \leq p \\ 1 - \hat{\alpha} & \text{if } j = p+1 \end{cases}.$$

From their definition, the estimators $\hat{\beta}_j$, for $j = 1, \dots, p$, are such that $\mathcal{U}_j(\hat{\beta}_j, \hat{\alpha}_j, \tau) =$

0. Thus, one can write :

$$\begin{aligned} \frac{1}{\sqrt{n}} \mathcal{U}_j(\beta_j, \alpha_j, \tau) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau (Z_i(s) - \bar{Z}(s)) \left(\hat{\alpha}_j \hat{\beta}_j^T Z_i(s) Y_i(s) ds - \alpha_j \beta_j^T Z_i(s) Y_i(s) ds \right) \\ &= \left[\int_0^\tau \frac{1}{n} \sum_{i=1}^n (Z_i(s) Z_i^T(s) - \bar{Z}(s) \bar{Z}^T(s)) Y_i(s) ds \right] \times \left[\sqrt{n}(\hat{\alpha}_j - \alpha_j) \hat{\beta}_j + \sqrt{n} \alpha_j (\hat{\beta}_j - \beta_j) \right]. \end{aligned}$$

With the notation

$$\hat{A}(\tau) = \int_0^\tau [S_2(s) - S_1^{\otimes 2}(s)/S_0(s)] ds,$$

we obtain

$$\frac{1}{\sqrt{n}} \mathcal{U}_j(\beta_j, \alpha_j, \tau) = \hat{A}(\tau) \left[\sqrt{n}(\hat{\alpha}_j - \alpha_j) \hat{\beta}_j + \alpha_j \sqrt{n}(\hat{\beta}_j - \beta_j) \right].$$

It follows that

$$\hat{\beta}_j - \beta_j = \frac{1}{\hat{\alpha}_j} \hat{A}^{-1}(\tau) \frac{1}{n} \mathcal{U}_j(\beta_j, \alpha_j, \tau) + \left(\frac{\alpha_j}{\hat{\alpha}_j} - 1 \right) \beta_j, \quad (6)$$

thanks to Assumption A4 which insures that $\hat{A}(\tau)$ is invertible. Now, using Assumptions A3 and A4, Proposition 4.1 and Lemma 4.2, it is easily seen that the right-hand side of (6) converges to zero when n tends to infinity. Thus $\hat{\beta}$ is consistent.

Furthermore, after some straightforward calculations on equation (6), one can write, for $j = 1, \dots, p+1$:

$$\begin{aligned} \sqrt{n}(\hat{\beta}_j - \beta_j) &= \frac{1}{\alpha_j} \left(A^{-1}(\tau) \frac{\mathcal{U}_j(\beta_j, \alpha_j, \tau)}{\sqrt{n}} - \beta_j \sqrt{n}(\hat{\alpha}_j - \alpha_j) \right) \\ &\quad + \frac{1}{\alpha_j} \left(\left(\hat{A}^{-1}(\tau) - A^{-1}(\tau) \right) \frac{\mathcal{U}_j(\beta_j, \alpha_j, \tau)}{\sqrt{n}} - \sqrt{n}(\hat{\alpha}_j - \alpha_j)(\hat{\beta}_j - \beta_j) \right). \end{aligned}$$

Using again Assumptions A3 and A4 (which ensure that $\hat{A}^{-1}(\tau)$ converges in probability to $A^{-1}(\tau)$), Proposition 4.1 and Lemma 4.2 as well as the consistency of $\hat{\beta}$, one can prove that the second term of the right-hand side of this last equality is an $o_P(1)$. Thus we can write:

$$\sqrt{n}(\hat{\beta}_j - \beta_j) = \frac{1}{\alpha} \left(A^{-1}(\tau) \frac{\mathcal{U}_j(\beta_j, \alpha_j, \tau)}{\sqrt{n}} - \beta_j \sqrt{n}(\hat{\alpha} - \alpha) \right) + o_P(1), \quad (7)$$

for $j = 1 \dots, p$, and

$$= \frac{1}{1-\alpha} \left(A^{-1}(\tau) \frac{\mathcal{U}_{p+1}(\beta_{p+1}, \alpha_{p+1}, \tau)}{\sqrt{n}} + \beta_{p+1} \sqrt{n}(\hat{\alpha} - \alpha) \right) + o_P(1). \quad (8)$$

Hence, from (4), (7) and (8) we finally obtain

$$\sqrt{n}(\hat{\beta} - \beta) = \Sigma_1(\tau) \mathbb{U}_n(\tau) + o_P(1), \quad (9)$$

where $\Sigma_1(\tau)$ is given in Theorem 4.3. This and Proposition 4.1 complete the proof. Note that because $\Sigma_{\mathbb{U}_\infty}(\tau)$ is positive definite, the matrix $\Sigma_{\hat{\beta}, \infty}$ is positive definite too. \square

4.4. The optimal estimator $\tilde{\beta}$

Recall that our T -optimal estimator is $\tilde{\beta} = \hat{H}\hat{\beta}$, where

$$\hat{H} = \operatorname{argmin}_{H \in \mathcal{H}} \hat{q}(H) = \operatorname{argmin}_{H \in \mathcal{H}} \operatorname{trace}(H \hat{\Sigma}_{\hat{\beta}, \infty} H^T),$$

and $\hat{\Sigma}_{\hat{\beta}, \infty}$ is required to be an estimator of the asymptotic covariance matrix $\Sigma_{\hat{\beta}, \infty}$ given in Theorem 4.3. Thus, we first have to find such an estimator of $\Sigma_{\hat{\beta}, \infty}$.

It is easy to check from their definition that the explicit expressions of the estimators $\hat{\beta}_j$, for $j = 1, \dots, p+1$, are

$$\hat{\beta}_j = \frac{1}{\hat{\alpha}_j} \hat{A}^{-1}(\tau) \frac{1}{n} \sum_{i=1}^n \int_0^\tau (Z_i(s) - \bar{Z}(s)) dN_{ij}(s). \quad (10)$$

Furthermore, let

$$\hat{\Theta}_j(\tau) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau (Z_i(s) - \bar{Z}(s))^{\otimes 2} dN_{ij}(s),$$

and

$$\hat{\xi}_j(\tau) = \kappa_j \frac{\hat{\alpha}(1 - \hat{\alpha})}{\hat{\theta}(\tau)} \hat{A}(\tau) \hat{\beta}_j,$$

where $\hat{\theta}(\tau) = N_{..}(\tau)/n$, and finally

$$\hat{\xi}_\alpha(\tau) = \frac{\hat{\alpha}(1 - \hat{\alpha})}{\hat{\theta}(\tau)}.$$

Therefore, one can use respectively

$$\hat{\Sigma}_1(\tau) = \begin{pmatrix} \frac{\hat{A}^{-1}(\tau)}{\hat{\alpha}} & 0 & \cdots & 0 & -\frac{\hat{\beta}_1}{\hat{\alpha}} \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \frac{\hat{A}^{-1}(\tau)}{\hat{\alpha}} & 0 & -\frac{\hat{\beta}_p}{\hat{\alpha}} \\ 0 & \cdots & 0 & \frac{\hat{A}^{-1}(\tau)}{1-\hat{\alpha}} & \frac{\hat{\beta}_{p+1}}{1-\hat{\alpha}} \end{pmatrix}$$

and

$$\hat{\Sigma}_{\mathbb{U}_\infty}(\tau) = \begin{pmatrix} \hat{\Theta}_1(\tau) & 0 & 0 & \hat{\xi}_1(\tau) \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & \hat{\Theta}_{p+1}(\tau) & \hat{\xi}_{p+1}(\tau) \\ \hat{\xi}_1^T(\tau) & \cdots & \hat{\xi}_{p+1}^T(\tau) & \hat{\xi}_\alpha(\tau) \end{pmatrix}$$

as estimators of $\Sigma_1(\tau)$ and $\Sigma_{\mathbb{U}_\infty}(\tau)$. Finally, let $\beta^* = (\beta_1^T, \dots, \beta_p^T)^T$.

Theorem 4.4. *Under Assumptions A1-A6, the estimator*

$$\hat{\Sigma}_{\hat{\beta}, \infty} = \hat{\Sigma}_1(\tau) \hat{\Sigma}_{\mathbb{U}_\infty}(\tau) \hat{\Sigma}_1^T(\tau)$$

converges in probability to the matrix $\Sigma_{\hat{\beta}, \infty}$. If $\Sigma_{\hat{\beta}, \infty}$ is invertible, then $\sqrt{n}(\tilde{\beta} - \beta^)$ is asymptotically zero mean gaussian distributed with covariance matrix whose trace minimizes $q(H) = \text{trace}(H \Sigma_{\hat{\beta}, \infty} H^T)$ over $H \in \mathcal{H}$.*

Proof. We know from Lemma 4.2 that $\hat{\alpha}$ converges in probability to α . We have also seen in the proof of Theorem 4.3 that $\hat{A}^{-1}(\tau)$ and $\hat{\beta}_j$, for $j = 1, \dots, p+1$, converge in probability respectively to $A^{-1}(\tau)$ and β_j . Thus, we get the convergence in probability of $\hat{\Sigma}_1(\tau)$ to $\Sigma_1(\tau)$.

Moreover, we have seen in the proof of Lemma 4.2 that

$$N_{\cdot}(\tau) = M_{\cdot}(\tau) + n \int_0^\tau [S_0(u) \lambda_{0p+1}(u) + \beta_{p+1}^T S_1(u)] du.$$

Thanks to Assumptions A3 and A4, the integral in the right-hand side of the last equation converges in probability to $\theta(\tau)$. This and (5) give the convergence in probability of $\hat{\theta}(\tau)$ to $\theta(\tau)$.

From the above convergences we have the convergence in probability of $\hat{\xi}_j(\tau)$ to $\xi_j(\tau)$, for $j = 1, \dots, p+1$, as well as the one of $\hat{\xi}_\alpha$ to ξ_α .

On the other hand, one can write, for $j = 1, \dots, p$:

$$\hat{\Theta}_j(\tau) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau (Z_i(s) - \bar{Z}(s))^{\otimes 2} dM_{ij}(s) + \frac{1}{n} \sum_{i=1}^n \int_0^\tau (Z_i(s) - \bar{Z}(s))^{\otimes 2} Y_i(s) \lambda'_j(s|Z_i) ds. \quad (11)$$

With the notation $A^{\otimes 2} = AA^T$ when A is also a matrix, it is easily seen that, for all $t \in [0, \tau]$

$$\left\langle \frac{1}{n} \sum_{i=1}^n \int_0^t (Z_i(s) - \bar{Z}(s))^{\otimes 2} dM_{ij}(s) \right\rangle = \frac{1}{n^2} \sum_{i=1}^n \int_0^t \left((Z_i(s) - \bar{Z}(s))^{\otimes 2} \right)^{\otimes 2} Y_i(s) \lambda'_j(s|Z_i) ds,$$

which in $t = \tau$ converges to zero in probability when n tends to infinity, by Assumptions A2 and A3. This and Lengart inequality yield the convergence in probability to zero of the first term of (11). Moreover, we have seen in the proof of Proposition 4.1 that the second term of (11) is equal to $\langle \mathcal{U}_j(\beta_j, \alpha, \tau) / \sqrt{n} \rangle$ and converges in probability to $\Theta_j(\tau)$, when n tends to infinity. These two last convergences prove that $\hat{\Theta}_j(\tau)$ converges to $\Theta_j(\tau)$, for $j = 1, \dots, p+1$. Thus we get the consistency of the estimator $\hat{\Sigma}_{\mathbb{U}_\infty}(\tau)$ and finally the one of $\hat{\Sigma}_{\hat{\beta}, \infty}$.

From the above and the continuous dependency of \hat{H} on $\hat{\Sigma}_{\hat{\beta}, \infty}$ we deduce that \hat{H} converges in probability to the matrix H_{opt} in \mathcal{H} which minimizes $\text{trace}(H \Sigma_{\hat{\beta}, \infty} H^T)$. We recall that the existence of such an optimal matrix is ensured because $\Sigma_{\hat{\beta}, \infty}$ is positive definite. Since $\hat{H}\beta = \beta^*$, we get from the above and Theorem 4.3 that $\sqrt{n}(\tilde{\beta} - \beta^*) = \hat{H}\sqrt{n}(\hat{\beta} - \beta)$ converges to a zero mean gaussian distributed random variable with covariance matrix $H_{opt} \Sigma_{\hat{\beta}, \infty} H_{opt}^T$ which is optimal in the sense defined earlier. \square

5. Statistical inference on the functional parameters

Even if it is not the model considered at the beginning of this paper, we will first consider the fully nonparametric case, that is a model without covariates where only functional parameters have to be estimated. Of course, this is also of interest for applications. Then we will come back to our semiparametric model and will see how to estimate its functional parameters.

5.1. Inference in the nonparameteric model

In this case no parametric form is assumed on the cause specific hazard rate functions $\lambda_j(\cdot)$, for $j = 1, \dots, p$ and we also do not take into account any covariate. Let $\Lambda_j(\cdot)$, for $j = 1, \dots, p$, denote the cause specific cumulative hazard rate functions defined by:

$$\Lambda_j(t) = \int_0^t \lambda_j(u) du,$$

for t in $[0, \tau]$ and

$$\Lambda_{p+1}(t) = \sum_{j=1}^p \Lambda_j(t).$$

With the assumption on M and the hypothesis of independence between (T, d) and C , the instantaneous transition rates of the Markov process with graph given in Figure 1 are:

$$\begin{cases} \lambda'_j(t) &= \alpha \lambda_j(t), \text{ for } j = 1, \dots, p, \\ \lambda'_{p+1}(t) &= (1 - \alpha) \sum_{j=1}^p \lambda_j(t) = (1 - \alpha) \lambda_{p+1}(t), \\ \lambda'_C(t) &= \lambda_C(t). \end{cases}$$

Contrary to the model by Goetghebeur and Ryan [14], when there is no covariate in the data, the model is still of interest because it allows different failure rates for failure causes.

5.1.1. First estimators of the cause specific cumulative hazard rate functions

By standard arguments, one can easily estimate $\Lambda_j(t)$, for $j = 1, \dots, p + 1$ and $t \in [0, \tau]$, by

$$\hat{\Lambda}_j(t) = \frac{1}{\hat{\alpha}_j} \hat{\Lambda}'_j(t),$$

where $\hat{\alpha}_j$ is defined at the beginning of the proof of Theorem 4.3, $Y(s) = \sum_{i=1}^n Y_i(s)$ is the number at risk at time s and

$$\hat{\Lambda}'_j(t) = \sum_{i=1}^n \int_0^t \frac{dN_{ij}(s)}{Y(s)}$$

is the well-known Nelson-Aalen estimator of

$$\Lambda'_j(t) = \int_0^t \lambda'_j(u) du.$$

Write:

$$\Lambda(\cdot) = \begin{pmatrix} \Lambda_1(\cdot) \\ \vdots \\ \Lambda_p(\cdot) \\ \Lambda_{p+1}(\cdot) \end{pmatrix} \text{ and } \hat{\Lambda}(\cdot) = \begin{pmatrix} \hat{\Lambda}_1(\cdot) \\ \vdots \\ \hat{\Lambda}_p(\cdot) \\ \hat{\Lambda}_{p+1}(\cdot) \end{pmatrix}.$$

Let us introduce two classical assumptions that allow to obtain the following asymptotic results.

B1. τ satisfies $0 < \Lambda_j(\tau) < +\infty$, for $j = 1, \dots, p$.

B2. There exists a function $s_0(\cdot)$, defined on $[0, \tau]$, and bounded away from 0, such that

$$\sup_{s \in [0, \tau]} \left| \frac{Y(s)}{n} - s_0(s) \right| \xrightarrow{P} 0, \text{ when } n \rightarrow +\infty.$$

Note that these assumptions are nothing but Assumptions A2 and A3 of Section 4 adapted to this new model. It is well known that Assumptions B1 and B2 are fulfilled whenever τ is such that $S(\tau)\bar{G}(\tau) > 0$, where $S(\cdot)$ and $\bar{G}(\cdot)$ are respectively the survival functions of T and C .

Note also that the functions $s_0(\cdot)\lambda_j(\cdot)$ and $\lambda_j(\cdot)/s_0(\cdot)$ are integrable on $[0, \tau]$.

Finally, let us define:

$$\begin{aligned} \eta_j(\cdot) &= \alpha_j \int_0^\cdot \lambda_j(u)/s_0(u) du, \text{ for } j = 1, \dots, p, \\ \eta_{p+1}(\cdot) &= \sum_{j=1}^p \eta_j(\cdot), \\ \theta_0(\cdot) &= \int_0^\cdot s_0(u)\lambda_m(u) du \\ \rho_j(\cdot) &= \kappa_j \alpha (1 - \alpha) \Lambda_j(\cdot)/\theta_0(\tau), \text{ for } 1 \leq j \leq p+1, \\ \text{and } \rho_\alpha(\tau) &= \alpha(1 - \alpha)/\theta_0(\tau). \end{aligned}$$

Theorem 5.1. *Under Assumptions A1, B1 and B2, the process $\sqrt{n}(\hat{\Lambda}(\cdot) - \Lambda(\cdot))$ converges weakly in $D^{p+1}[0, \tau]$ to a zero mean gaussian process $\mathbb{L}(\cdot)$ with covariance matrix function*

$$\Sigma_{\mathbb{L}}(t) = \Sigma_2(t) \Sigma_{\mathbb{V}_\infty}(t) \Sigma_2^T(t),$$

where

$$\Sigma_2(t) = \begin{pmatrix} \frac{1}{\alpha} & 0 & \cdots & 0 & -\frac{\Lambda_1(t)}{\alpha} \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \frac{1}{\alpha} & 0 & -\frac{\Lambda_p(t)}{\alpha} \\ 0 & \cdots & 0 & \frac{1}{1-\alpha} & \frac{\Lambda_{p+1}(t)}{1-\alpha} \end{pmatrix}$$

and

$$\Sigma_{\mathbb{V}_\infty}(t) = \begin{pmatrix} \eta_1(t) & 0 & 0 & \rho_1(t) \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & \eta_{p+1}(t) & \rho_{p+1}(t) \\ \rho_1(t) & \cdots & \rho_{p+1}(t) & \rho_\alpha(\tau) \end{pmatrix}.$$

Proof. Our proof starts with the observation that:

$$\hat{\Lambda}_j(t) - \Lambda_j(t) = \frac{1}{\hat{\alpha}_j} \left(\hat{\Lambda}'_j(t) - \Lambda'_j(t) \right) - \frac{\Lambda'_j(t)}{\sqrt{n}\alpha_j\alpha'_j} \sqrt{n}(\hat{\alpha}_j - \alpha). \quad (12)$$

A straightforward application of the central limit theorem gives the \sqrt{n} -asymptotical normality of $\hat{\alpha}$ and thus that $\sqrt{n}(\hat{\alpha} - \alpha)$ is an $O_p(1)$. Moreover, under Assumptions B1 and B2, the Nelson-Aalen estimators $\hat{\Lambda}'_j(\cdot)$, for $j = 1, \dots, p+1$, are well-known to be uniformly consistent (see *e.g.* Andersen *et al.* [1]). Hence, under the assumptions of Theorem 5.1, one can get easily from (12) the uniform convergence in $D[0, \tau]$ of $\hat{\Lambda}_j(\cdot)$ to $\Lambda_j(\cdot)$, for $j = 1, \dots, p+1$. These convergences and (12) yield

$$\sqrt{n}(\hat{\Lambda}_j(\cdot) - \Lambda_j(\cdot)) = \frac{\sqrt{n}}{\alpha_j} \int_0^\cdot \frac{dM_{\cdot j}(s)}{Y(s)} - \frac{\Lambda_j(\cdot)}{\alpha_j} \sqrt{n}(\hat{\alpha}_j - \alpha_j) + o_P(1), \quad (13)$$

in $D[0, \tau]$ and for $j = 1, \dots, p+1$.

But, following the lines of Lemma 4.2, we show that

$$\sqrt{n}(\hat{\alpha} - \alpha) = \sum_{j=1}^{p+1} \frac{\alpha_j^*}{\theta_0(\tau)} \frac{1}{\sqrt{n}} M_{\cdot j}(\tau) + o_P(1),$$

where $\theta_0(\tau) > 0$ by B1 and B2. From this and (13), we deduce that

$$\sqrt{n}(\hat{\Lambda}(t) - \Lambda(t)) = \Sigma_2(t) \mathbb{V}_n(t) + o_P(1), \quad (14)$$

in $D^{p+1}[0, \tau]$, where $\Sigma_2(t)$ is defined in Theorem 5.1 and

$$\mathbb{V}_n(t) = \sqrt{n} \begin{pmatrix} \int_0^t \frac{dM_{\cdot 1}(s)}{Y(s)} \\ \vdots \\ \int_0^t \frac{dM_{\cdot p+1}(s)}{Y(s)} \\ \frac{1}{n} \sum_{j=1}^{p+1} \frac{\alpha_j^*}{\theta_0(\tau)} M_{\cdot j}(\tau) \end{pmatrix}.$$

Using the Rebolledo Theorem or an approach given in Dauxois [8] one can easily prove that, under Assumptions B1 and B2, the process $\mathbb{V}_n(\cdot)$ converges weakly in $D^{p+2}[0, \tau]$ to a zero mean gaussian process $\mathbb{V}_\infty(\cdot)$. It remains to specify its covariance function $\Sigma_{\mathbb{V}_\infty}(\cdot)$. First, note that we can write $\mathbb{V}_n(t) = \mathbb{V}_{1,n}(t) + \mathbb{V}_{2,n}(t)$ where

$$\mathbb{V}_{1,n}(t) = \sqrt{n} \begin{pmatrix} \int_0^t \frac{dM_{\cdot 1}(s)}{Y(s)} \\ \vdots \\ \int_0^t \frac{dM_{\cdot p+1}(s)}{Y(s)} \\ \frac{1}{n} \sum_{j=1}^{p+1} \frac{\alpha_j^*}{\theta_0(\tau)} M_{\cdot j}(t) \end{pmatrix}$$

and

$$\mathbb{V}_{2,n}(t) = \frac{1}{\sqrt{n}} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \sum_{j=1}^{p+1} \frac{\alpha_j^*}{\theta_0(\tau)} (M_{\cdot j}(\tau) - M_{\cdot j}(t)) \end{pmatrix}.$$

Since the martingales $M_{\cdot j}$ are orthogonal, for $j = 1, \dots, p+1$, we have

$$\mathbb{E} [\mathbb{V}_n^{\otimes 2}(t)] = \mathbb{E} [\mathbb{V}_{1,n}^{\otimes 2}(t)] + \mathbb{E} [\mathbb{V}_{2,n}^{\otimes 2}(t)]$$

and thus

$$\Sigma_{\mathbb{V}_\infty}(t) = \lim_{n \rightarrow \infty} \mathbb{E} [\mathbb{V}_{1,n}^{\otimes 2}(t)] + \lim_{n \rightarrow \infty} \mathbb{E} [\mathbb{V}_{2,n}^{\otimes 2}(t)].$$

Straightforward calculus lead to the expression of $\Sigma_{\mathbb{V}_\infty}(\cdot)$ given in the theorem.

From this and (14) we get the desired result. \square

One can get a consistent estimator of the covariance function $\Sigma_{\mathbb{L}}(\cdot)$ of the limit process $\mathbb{L}(\cdot)$. In this order, let us define, for $j = 1, \dots, p+1$:

$$\hat{\eta}_j(t) = n \int_0^t \frac{dN_{\cdot j}(s)}{Y^2(s)}.$$

Define also:

$$\hat{\rho}_j(t) = \kappa_j \frac{\hat{\alpha}(1 - \hat{\alpha})\hat{\Lambda}_j(t)}{\hat{\theta}_0(\tau)}, \text{ for } j = 1, \dots, p, \text{ and } \hat{\rho}_\alpha = \frac{\hat{\alpha}(1 - \hat{\alpha})}{\hat{\theta}_0(\tau)},$$

where $\hat{\theta}_0(\tau) = N_{..}(\tau)/n$.

Then, plug-in estimators of $\Sigma_2(t)$ and $\Sigma_{\mathbb{V}_\infty}(t)$, for $t \in [0, \tau]$, are respectively

$$\hat{\Sigma}_2(t) = \begin{pmatrix} \frac{1}{\hat{\alpha}} & 0 & \cdots & 0 & -\frac{\hat{\Lambda}_1(t)}{\hat{\alpha}} \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \frac{1}{\hat{\alpha}} & 0 & -\frac{\hat{\Lambda}_p(t)}{\hat{\alpha}} \\ 0 & \cdots & 0 & \frac{1}{1-\hat{\alpha}} & \frac{\hat{\Lambda}_{p+1}(t)}{1-\hat{\alpha}} \end{pmatrix}$$

and

$$\hat{\Sigma}_{\mathbb{V}_\infty}(t) = \begin{pmatrix} \hat{\eta}_1(t) & 0 & 0 & \hat{\rho}_1(t) \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & \hat{\eta}_{p+1}(t) & \hat{\rho}_{p+1}(t) \\ \hat{\rho}_1(t) & \cdots & \hat{\rho}_{p+1}(t) & \hat{\rho}_\alpha \end{pmatrix}.$$

Finally, an empirical estimator of $\Sigma_{\mathbb{L}}(\cdot)$ is given by:

$$\hat{\Sigma}_{\mathbb{L}}(\cdot) = \hat{\Sigma}_2(\cdot)\hat{\Sigma}_{\mathbb{V}_\infty}(\cdot)\hat{\Sigma}_2^T(\cdot).$$

5.1.2. Optimal estimators of the cause specific cumulative hazard rate functions

Except the fact that estimators in this section are functions the situation is the same as the one of Section 3.1. Indeed, our multivariate functional estimator $\hat{\Lambda}(\cdot)$ gives estimators $\hat{\Lambda}_1(\cdot), \dots, \hat{\Lambda}_p(\cdot)$ of $\Lambda_1(\cdot), \dots, \Lambda_p(\cdot)$ respectively and an estimator $\hat{\Lambda}_{p+1}(\cdot)$ of their sum $\Lambda_{p+1}(\cdot) = \sum_{i=1}^p \Lambda_i(\cdot)$. Here also we will look for a linear transformation of $\hat{\Lambda}(\cdot)$ which will give us an optimal estimator of $\Lambda^*(\cdot) = (\Lambda_1(\cdot), \dots, \Lambda_p(\cdot))^T$.

In this order, let us define \mathcal{H}' as the set of $p \times (p+1)$ real valued matrices such that $Ha = a^*$ for all $a^* = (a_1, \dots, a_p)^T \in \mathbb{R}^p$ and $a = (a^{*T}, \sum_{j=1}^p a_j)^T$. We define

$$\hat{H}(t) = \underset{H \in \mathcal{H}'}{\operatorname{argmin}} \operatorname{trace}(H\hat{\Sigma}_{\mathbb{L}}(t)H^T) \quad (15)$$

and set $\tilde{\Lambda}(\cdot) = \hat{H}(\cdot)\hat{\Lambda}(\cdot)$ as a new estimator of $\Lambda^*(\cdot)$. The next theorem proves that the later estimator is asymptotically normal and T -optimal.

Theorem 5.2. *Under Assumptions A1, B1 and B2, the matrix $\hat{\Sigma}_{\mathbb{L}}(t)$ converges in probability to the matrix $\Sigma_{\mathbb{L}}(t)$, for all $t \in [0, \tau]$.*

Moreover, let us assume that the matrix $\Sigma_{\mathbb{L}}(t)$ is invertible, for all $t \in]0, \tau]$. If, for all $t \in [0, \tau]$, the matrix $\hat{H}(t)$ is the unique solution of (15), then the process $\sqrt{n}(\tilde{\Lambda}(\cdot) - \Lambda^(\cdot))$ converges weakly in $D^p[0, \tau]$ to a centered gaussian process $\mathbb{L}'(\cdot) = H_{\Sigma_{\mathbb{L}}}(\cdot)\mathbb{L}(\cdot)$ with covariance function $H_{\Sigma_{\mathbb{L}}}(\cdot)\Sigma_{\mathbb{L}}(\cdot)H_{\Sigma_{\mathbb{L}}}^T(\cdot)$ where*

$$H_{\Sigma_{\mathbb{L}}}(t) = \underset{H \in \mathcal{H}'}{\operatorname{argmin}} \operatorname{trace}(H\Sigma_{\mathbb{L}}(t)H^T).$$

Proof. The proof is omitted since it follows the lines of the proof of Theorem 4.4. \square

Let us denote by L the column vector in $\mathbb{R}^{p(p+1)}$ defined by $L = (H_1, \dots, H_p)$ where H_i is the i th row of $H \in \mathcal{H}'$. The link between $L = (l_i)_{1 \leq i \leq p(p+1)}$ and $H = (h_{i,j})_{1 \leq i \leq p; 1 \leq j \leq p+1}$ is therefore $h_{i,j} = l_{(i-1)(p+1)+j}$. One can see that the linear constraints on H may be written on L as $CL = d$ where C and d are known. Indeed

$$\begin{aligned} & \begin{cases} h_{i,i} + h_{i,p+1} &= 1, \text{ for } 1 \leq i \leq p, \\ h_{i,j} + h_{i,p+1} &= 0, \text{ for } 1 \leq i \leq p, 1 \leq j \leq p+1 \text{ and } i \neq j, \end{cases} \\ \Leftrightarrow & \begin{cases} l_{(i-1)(p+1)+i} + l_{i(p+1)} &= 1, \text{ for } 1 \leq i \leq p, \\ l_{(i-1)(p+1)+j} + l_{(i+1)p} &= 0, \text{ for } 1 \leq i \leq p, 1 \leq j \leq p+1 \text{ and } i \neq j, \end{cases} \\ \Leftrightarrow & CL = d, \end{aligned}$$

where the matrix C and the vector d are obvious. Moreover, let $\hat{Q}(t)$ be the $p \times p$ block diagonal matrix defined by

$$\hat{Q}(t) = \operatorname{diag} \left(\underbrace{\hat{\Sigma}_{\mathbb{L}}(t), \dots, \hat{\Sigma}_{\mathbb{L}}(t)}_{p \text{ times}} \right),$$

and note that

$$\operatorname{trace}(H\hat{\Sigma}_{\mathbb{L}}(t)H^T) = L^T \hat{Q}(t)L.$$

Thus, in order to find our optimal estimator, we have to solve the following optimization problems

$$(P_t) \quad \begin{cases} \text{Find } L(t) \text{ which minimizes } L^T(t)\hat{Q}(t)L(t) \\ \text{such that : } CL(t) = d \end{cases},$$

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

Following the method of Section 3.2, the solution of (P_t) is:

$$\hat{L}(t) = \hat{Q}^{-1}(t)C^T(C\hat{Q}^{-1}(t)C^T)^{-1}d$$

and $\hat{H}(t)$ is therefore defined by $\hat{h}_{ij}(t) = \hat{l}_{(i-1)(p+1)+j}(t)$ for $1 \leq i \leq p$ and $1 \leq j \leq p+1$.

Remark. Since $\hat{\Sigma}_{\mathbb{L}}(t)$ is piecewise constant, it is sufficient to calculate the matrix $\hat{H}(t)$ at points $t \in [0, \tau]$ where the counting process N_{\cdot} jumps, that is at points $X_i \in [0, \tau]$ such that $\delta_i = 1$.

5.1.3. Estimation of the cumulative incidence functions and the survival function of T

Our aim in this section is to introduce estimators of the survival function $S(\cdot)$ of the lifetime T as well as estimators of the cumulative incidence functions $F_j(\cdot)$ defined, for all t and $j = 1, \dots, p$, by

$$F_j(t) = P(T \leq t, d = j).$$

Let us recall that $\Lambda(\cdot) = \sum_{j=1}^p \Lambda_j(\cdot)$ is the cumulative hazard rate function of the survival time T . It is well known that one can write the survival function in terms of the cumulative hazard rate function:

$$S(t) = \prod_{u \in]0, t]} (1 - d\Lambda(u)),$$

where \prod denotes the product integral (see Gill & Johansen [13]). Using a plug-in method on this last equation, one can get the Kaplan-Meier estimator of $S(t)$:

$$\hat{S}(t) = \prod_{u \in]0, t]} (1 - d\tilde{\Lambda}(u)),$$

where $\tilde{\Lambda}(\cdot)$ is the optimal estimator of the previous paragraph and

$$\tilde{\Lambda}(\cdot) = \sum_{j=1}^p \tilde{\Lambda}_j(\cdot).$$

On the other hand, it is also well-known that one can write, for $t \in [0, \tau]$ and $j = 1, \dots, p$

$$F_j(t) = \int_0^t S(u^-) d\Lambda_j(u),$$

and that an estimator of this cumulative incidence function is given by the Aalen-Johansen estimator (see Andersen *et al.* [1])

$$\hat{F}_j(t) = \int_0^t \hat{S}(u^-) d\tilde{\Lambda}_j(u).$$

The asymptotic behaviour of these estimators is given in the following theorem.

Theorem 5.3. *Under Assumptions A1, B1 and B2, we have the following weak convergence in $D^{p+1}[0, \tau]$, when n tends to infinity*

$$\sqrt{n} \left[\begin{pmatrix} \hat{F}_1(\cdot) \\ \vdots \\ \hat{F}_p(\cdot) \\ \hat{S}(\cdot) \end{pmatrix} - \begin{pmatrix} F_1(\cdot) \\ \vdots \\ F_p(\cdot) \\ S(\cdot) \end{pmatrix} \right] \rightarrow \begin{pmatrix} \int_0^\cdot S(u) d\mathbb{L}'_1(u) - \int_0^\cdot S(u) \mathbb{L}'(u) d\Lambda_1(u) \\ \vdots \\ \int_0^\cdot S(u) d\mathbb{L}'_p(u) - \int_0^\cdot S(u) \mathbb{L}'(u) d\Lambda_p(u) \\ -S(\cdot) \mathbb{L}'(\cdot) \end{pmatrix},$$

where $\mathbb{L}'_j(\cdot)$ is the j th component of the limit process $\mathbb{L}'(\cdot)$ of Theorem 5.2 and $\mathbb{L}'(\cdot) = \mathbb{L}'_1(\cdot) + \dots + \mathbb{L}'_p(\cdot)$.

Proof. This result is easily obtained from Theorem 5.2 and the functional δ -method (see *e.g.* van der Vaart & Wellner [27] for details on this method). Indeed, from the above, one can write

$$\begin{pmatrix} F_1(\cdot) \\ \vdots \\ F_p(\cdot) \\ S(\cdot) \end{pmatrix} = \psi \begin{pmatrix} \Lambda_1(\cdot) \\ \vdots \\ \Lambda_p(\cdot) \end{pmatrix},$$

where $\psi(\cdot)$ is a function from $D^p[0, \tau]$ to $D^{p+1}[0, \tau]$ with coordinate functions defined, for $i = 1, \dots, p$, by:

$$F_i(\cdot) = \int_0^\cdot S(u^-) d\Lambda_i(u) = \psi_i \begin{pmatrix} \Lambda_1(\cdot) \\ \vdots \\ \Lambda_p(\cdot) \end{pmatrix},$$

and

$$S(\cdot) = \psi_{p+1} \begin{pmatrix} \Lambda_1(\cdot) \\ \vdots \\ \Lambda_p(\cdot) \end{pmatrix} = \prod_{u \in]0, \cdot]} (1 - d\Lambda(u)).$$

We know (see again [27]) that the product-integral function

$$\begin{aligned} \phi(\cdot) : D[0, \tau] &\rightarrow D[0, \tau] \\ A &\mapsto \phi(A) = \prod_{u \in]0, \cdot]} (1 - dA(u)) \end{aligned}$$

is Hadamard differentiable on $BV_K[0, \tau]$, the subset of functions in $D[0, \tau]$ with total variation bounded by K . Its derivative is

$$\phi'_A(\alpha)(\cdot) = - \int_0^\cdot \phi(A)(u^-) \frac{\phi(A)(\cdot)}{\phi(A)(u)} d\alpha(u)$$

and is equal to $-\phi(A)(\cdot)\alpha(\cdot)$ when A is continuous.

We also know that the function

$$\begin{aligned} \varphi(\cdot, \cdot) : D[0, \tau] \times BV_K[0, \tau] &\rightarrow D[0, \tau] \\ (A, B) &\mapsto \varphi(A, B) = \int_0^\cdot A dB \end{aligned}$$

is Hadamard differentiable on (A, B) where A is of bounded variation, with derivative:

$$\varphi'_{A,B}(\alpha, \beta)(\cdot) = \int_0^\cdot A(u) d\beta(u) + \int_0^\cdot \alpha(u) dB(u).$$

Using for instance the chain rule (see Lemma 3.9.3 of [27]) and the composition

$$(\Lambda_1(\cdot), \dots, \Lambda_p(\cdot))^T \mapsto (\Lambda_i(\cdot), S(\cdot))^T \mapsto \int_0^\cdot S(u^-) d\Lambda_i(u),$$

one can prove that the coordinate functions $\psi_1(\cdot), \dots, \psi_p(\cdot)$ are differentiable and get their derivatives. The functional δ -method allows us to obtain from

Theorem 5.2 the following weak convergence in $D^{p+1}[0, \tau]$:

$$\sqrt{n} \left(\psi(\tilde{\Lambda}(\cdot)) - \psi(\Lambda^*(\cdot)) \right) \rightarrow \psi'_{\Lambda^*}(\mathbb{L}'(\cdot)),$$

when n tends to infinity. Straightforward differential calculus and the continuity of $S(\cdot)$ give the expression of $\psi'_{\Lambda^*}(\mathbb{L}'(\cdot))$ detailed in the Theorem. \square

5.2. With explanatory variables

Now, let us come back to the semiparametric model of equation (1) with explanatory variable $Z(\cdot)$. Recall that we have, for $j = 1, \dots, p$:

$$dN_{\cdot j}(s) = dM_{\cdot j}(s) + \alpha Y(s) d\Lambda_{0j}(s) + \alpha n S_1^T(s) \beta_j ds.$$

Thus, an estimator of $\Lambda_{0j}(t)$ is given, for all $t \in [0, \tau]$, by

$$\hat{\Lambda}_{0j}(t) = \frac{1}{\hat{\alpha}} \int_0^t \frac{dN_{\cdot j}(s)}{Y(s)} - \hat{\beta}_j^T \int_0^t \frac{S_1(s)}{S_0(s)} ds,$$

where $\hat{\beta}_j$ is the estimator introduced in Section 3.1 and is such that $\mathcal{U}_j(\hat{\beta}_j, \alpha_j, \tau) = 0$.

Then we propose to estimate $\Lambda_j(\cdot|Z)$, for $j = 1, \dots, p$, by $\hat{\Lambda}_j(\cdot|Z)$ defined, for $t \geq 0$, by

$$\hat{\Lambda}_j(t|Z) = \hat{\Lambda}_{0j}(t) + \hat{\beta}_j^T \int_0^t Z(s) ds = \frac{1}{\hat{\alpha}} \int_0^t \frac{dN_{\cdot j}(s)}{Y(s)} + \hat{\beta}_j^T \left(\int_0^t Z(s) ds - \int_0^t \frac{S_1(s)}{S_0(s)} ds \right),$$

where by abuse of notation $\int_0^t Z(s) ds$ denotes the vector of integrals of each coordinate of $Z(\cdot)$.

Before giving the large sample behaviour of these estimators in the following theorem, let us introduce other notations. For $j = 1, \dots, p$ let us define:

$$\begin{aligned} \mu_{1j}(\cdot) &= \alpha \int_0^\cdot \frac{s_0(s) \lambda_{0j}(s) + \beta_j^T s_1(s)}{s_0^2(s)} ds, \\ \mu_{2j}(\cdot) &= \frac{\alpha(1-\alpha)}{\theta(\tau)} \int_0^\cdot \frac{s_0(s) \lambda_{0j}(s) + \beta_j^T s_1(s)}{s_0(s)} ds, \\ \mu_{3j}(\cdot) &= \alpha \int_0^\cdot \frac{a(s)}{s_0(s)} ds \beta_j, \end{aligned}$$

and the block matrix

$$\Sigma_5(t) = \frac{1}{\alpha} \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & K(t) & 0 & 0 & \dots & 0 & -\Lambda_1(t|\mathbb{Z}) \\ 0 & 1 & 0 & \dots & 0 & 0 & K(t) & 0 & \dots & 0 & -\Lambda_2(t|\mathbb{Z}) \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 1 & 0 & 0 & \dots & 0 & K(t) & 0 & -\Lambda_{p-1}(t|\mathbb{Z}) \\ 0 & \dots & 0 & 0 & 1 & 0 & \dots & 0 & 0 & K(t) & -\Lambda_p(t|\mathbb{Z}) \end{pmatrix}$$

where $K(t)$ is a $1 \times k$ vector defined by

$$K(t) = \left(\int_0^t Z(s)ds - \int_0^t \frac{s_1(s)}{s_0(s)} ds \right)^T A^{-1}(\tau), \text{ for all } t \in [0, \tau].$$

Theorem 5.4. *Under Assumptions A1-A6, the process*

$$\sqrt{n} \left(\hat{\Lambda}_1(\cdot|\mathbb{Z}) - \Lambda_1(\cdot|\mathbb{Z}), \dots, \hat{\Lambda}_p(\cdot|\mathbb{Z}) - \Lambda_p(\cdot|\mathbb{Z}) \right)^T$$

converges weakly in $D^p[0, \tau]$, when $n \rightarrow +\infty$, to a zero mean gaussian process

with covariance matrix $\Sigma_5(t)\Sigma_{\mathbb{W}'_\infty}(t)\Sigma_5^T(t)$, where

$$\Sigma_{\mathbb{W}'_\infty}(t) = \frac{1}{\alpha} \begin{pmatrix} \Sigma_{\mathbb{W}'_\infty}^{(11)}(t) & \Sigma_{\mathbb{W}'_\infty}^{(12)}(t) \\ \Sigma_{\mathbb{W}'_\infty}^{(12)}(t)^T & \Sigma_{\mathbb{W}'_\infty}^{(22)}(t) \end{pmatrix},$$

with

$$\Sigma_{\mathbb{W}'_\infty}^{(11)}(t) = \text{diag} \underbrace{(\mu_{11}(t), \dots, \mu_{1p}(t))}_{p \text{ times}},$$

$$\Sigma_{\mathbb{W}'_\infty}^{(12)}(t) = \begin{pmatrix} \mu_{31}^T(t) & 0 & 0 & \dots & 0 & \mu_{21}(t) \\ 0 & \mu_{32}^T(t) & 0 & \dots & 0 & \mu_{22}(t) \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & \mu_{3(p-1)}^T(t) & 0 & \mu_{2(p-1)}(t) \\ 0 & \dots & 0 & 0 & \mu_{3p}^T(t) & \mu_{2p}(t) \end{pmatrix},$$

and

$$\Sigma_{\mathbb{W}'_\infty}^{(22)}(t) = \begin{pmatrix} \Theta_1(\tau) & 0 & 0 & \dots & 0 & \xi_1(\tau) \\ 0 & \Theta_2(\tau) & 0 & \dots & 0 & \xi_2(\tau) \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & \Theta_{p-1}(\tau) & 0 & \xi_{p-1}(\tau) \\ 0 & \dots & 0 & 0 & \Theta_p(\tau) & \xi_p(\tau) \\ \xi_1^T(\tau) & \dots & \dots & \dots & \xi_p^T(\tau) & \xi_\alpha(\tau) \end{pmatrix}.$$

Proof. By the definition of $\Lambda_j(\cdot|Z)$ and $\hat{\Lambda}_j(\cdot|Z)$, one can write, for all $t > 0$ and $j = 1, \dots, p$

$$\begin{aligned} & \sqrt{n} \left(\hat{\Lambda}_j(\cdot|Z) - \Lambda_j(\cdot|Z) \right) \\ = & \sqrt{n} \left[\left(\frac{1}{\hat{\alpha}} \int_0^t \frac{dN_{\cdot j}(s)}{Y(s)} - \hat{\beta}_j^T \int_0^t \frac{S_1(s)}{S_0(s)} ds - \Lambda_{0j}(t) \right) + (\hat{\beta}_j - \beta_j)^T \int_0^t Z(s) ds \right]. \end{aligned}$$

But it is easy to check that our assumptions and preliminary results yield

$$\begin{aligned} & \sqrt{n} \left(\frac{1}{\hat{\alpha}} \int_0^\cdot \frac{dN_{\cdot j}(s)}{Y(s)} - \hat{\beta}_j^T \int_0^\cdot \frac{S_1(s)}{S_0(s)} ds - \Lambda_{0j}(\cdot) \right) = \frac{\sqrt{n}}{\hat{\alpha}} \int_0^\cdot \frac{dM_{\cdot j}(s)}{Y(s)} \\ - & \frac{\sqrt{n}(\hat{\alpha} - \alpha)}{\alpha} \left(\beta_j^T \int_0^\cdot \frac{s_1(s)}{s_0(s)} ds + \Lambda_{0j}(\cdot) \right) - \sqrt{n}(\hat{\beta}_j - \beta_j)^T \int_0^\cdot \frac{s_1(s)}{s_0(s)} ds + o_P(1). \end{aligned}$$

These two last equations and (7) enable us to write:

$$\begin{aligned} & \sqrt{n} \left(\hat{\Lambda}_j(\cdot|Z) - \Lambda_j(\cdot|Z) \right) = \frac{\sqrt{n}}{\hat{\alpha}} \int_0^\cdot \frac{dM_{\cdot j}(s)}{Y(s)} - \frac{\sqrt{n}(\hat{\alpha} - \alpha)}{\alpha} \Lambda_j(\cdot|Z) \\ + & \frac{1}{\alpha} \left(A^{-1}(\tau) \frac{\mathcal{U}_j(\beta_j, \alpha_j, \tau)}{\sqrt{n}} \right)^T \left(\int_0^\cdot Z(s) ds - \int_0^\cdot \frac{s_1(s)}{s_0(s)} ds \right) + o_P(1). \end{aligned}$$

Hence we have

$$\sqrt{n} \begin{pmatrix} \hat{\Lambda}_1(\cdot|Z) - \Lambda_1(\cdot|Z) \\ \vdots \\ \hat{\Lambda}_p(\cdot|Z) - \Lambda_p(\cdot|Z) \end{pmatrix} = \Sigma_5(\cdot) \mathbb{W}'_n(\cdot) + o_P(1),$$

where

$$\begin{aligned} \mathbb{W}'_n(\cdot) = & \left(\sqrt{n} \int_0^\cdot \frac{dM_{\cdot 1}(s)}{Y(s)}, \dots, \sqrt{n} \int_0^\cdot \frac{dM_{\cdot p}(s)}{Y(s)}, \right. \\ & \left. \frac{\mathcal{U}_1^T(\beta_1, \alpha_1, \tau)}{\sqrt{n}}, \dots, \frac{\mathcal{U}_p^T(\beta_p, \alpha_p, \tau)}{\sqrt{n}}, \frac{1}{\sqrt{n}} \sum_{j=1}^{p+1} \frac{\alpha_j^*}{\theta(\tau)} M_{\cdot j}(\tau) \right)^T. \end{aligned}$$

But a straightforward application of Rebolledo's Theorem proves that, under Assumptions A1-A6, the process

$$\mathbb{W}_n(\cdot) = \left(\sqrt{n} \int_0^t \frac{dM_{\cdot 1}(s)}{Y(s)}, \dots, \sqrt{n} \int_0^t \frac{dM_{\cdot p}(s)}{Y(s)}, \right. \\ \left. \frac{\mathcal{U}_1(\beta_1, \alpha_1, \cdot)}{\sqrt{n}}, \dots, \frac{\mathcal{U}_p(\beta_p, \alpha_p, \cdot)}{\sqrt{n}}, \frac{1}{\sqrt{n}} \sum_{j=1}^{p+1} \frac{\alpha_j^*}{\theta(\tau)} M_{\cdot j}(\cdot) \right)^T$$

converges weakly in $D^{p(k+1)+1}[0, \tau]$ to a zero mean multivariate gaussian martingale $\mathbb{W}_\infty(\cdot)$. It is easy to check that the covariance matrix $\Sigma_{\mathbb{W}_\infty}(\cdot)$ of this limit process is similar to $\Sigma_{\mathbb{W}_\infty}(\cdot)$ where τ is everywhere replaced by t . This ends the proof. \square

Estimators $\hat{\Theta}_1(\tau), \dots, \hat{\Theta}_p(\tau), \hat{\xi}_1(\tau), \dots, \hat{\xi}_p(\tau)$ and $\hat{\xi}_\alpha(\tau)$ of respectively $\Theta_1(\tau), \dots, \Theta_p(\tau), \xi_1(\tau), \dots, \xi_p(\tau)$ and $\xi_\alpha(\tau)$ have been introduced in Section 4.4. Moreover, one can estimate $\mu_{1j}(t), \mu_{2j}(t), \mu_{3j}(t)$, for $j = 1, \dots, p$ and $t \in [0, \tau]$, with

$$\begin{aligned} \hat{\mu}_{1j}(t) &= n\hat{\alpha} \int_0^t \frac{dN_{\cdot j}}{S_0^2(s)}, \\ \hat{\mu}_{2j}(t) &= \frac{\hat{\alpha}(1 - \hat{\alpha})}{\hat{\theta}(\tau)} \int_0^t \frac{dN_{\cdot j}(s)}{S_0(s)}, \\ \hat{\mu}_{3j}(t) &= \hat{\alpha} \hat{A}(t) \hat{\beta}_j \end{aligned}$$

and an estimator of $K(t)$ is given by:

$$\hat{K}(t) = \left(\int_0^t Z(s) ds - \int_0^t \frac{S_1(s)}{S_0(s)} ds \right)^T \hat{A}^{-1}(\tau).$$

Using these estimators one can get consistent estimators of the matrices $\Sigma_5(t)$ and $\Sigma_{\mathbb{W}'_\infty}(t)$.

Finally, it has to be noted that, as in Section 5.1.3, it is also possible to get estimators of the survival function and the cumulative incidence functions as well as to get their large sample behaviour (like in Theorem 5.3). But this is omitted here since it is very similar to what has been done in Section 5.1.3.

6. Simulation study

A simulation study is driven in this section in order to assess the behavior of our estimates on small sample sizes. We have assumed that there is two compet-

ing risks and a single time-independent covariate Z with uniform distribution on $\{1, 2\}$. Moreover, the two risks T_1 and T_2 are supposed to be, conditionally on Z , independent with respective hazard rate function $\lambda_1(t|Z) = t/2 + 2Z$ and $\lambda_2(t|Z) = 2t/9 + 3Z$. Therefore we have $\beta_1 = 2$ and $\beta_2 = 3$. The censoring time C is supposed to be exponentially distributed with mean 1 whereas the rate of missingness α is equal to 0.5. As a consequence, the rate of censored data is about 12%, the rates of observed failures from cause 1 and 2 are respectively about 15% and 28%, and the rate of missingness is about 45%. The study of the performance of our estimators is based on 1000 simulations of samples with size 100, 200, 400 or 1000.

n	$\hat{\beta}_1$	$\tilde{\beta}_1$	$\hat{\beta}_2$	$\tilde{\beta}_2$
100	2.116 [1.583]	2.117 [1.438]	3.024 [1.923]	3.004 [1.610]
200	2.055 [1.111]	2.023 [1.000]	3.083 [1.367]	3.039 [1.146]
400	2.021 [0.768]	2.001 [0.687]	3.043 [0.919]	3.011 [0.784]
1000	1.998 [0.500]	2.001 [0.454]	2.989 [0.584]	2.993 [0.502]

Table 1: Empirical means [standard errors] of $\hat{\beta}_i$ and $\tilde{\beta}_i$, for $i = 1, 2$, based on Monte Carlo simulation of size $N=1000$.

Table 1 gives the mean and the standard error of the estimators $(\hat{\beta}_i)_{i=1,2}$, based on transitions $0 \rightarrow i$, and of the T -optimal estimators $(\tilde{\beta}_i)_{i=1,2}$, which use observations coming from the three informative transitions. It appears that in terms of bias the T -optimal estimators are generally better (except for β_1 when $n = 100$) whereas the standard errors of the T -optimal estimates are always smaller than the standard errors of estimates based on transitions 1 and 2.

References

- [1] Andersen, P.K., Borgan, O. Gill, R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer, New York.
- [2] Andersen, P.K. and Gill, R.D. (1982). Cox's regression model for counting processes: A large sample study, *Ann. Statist.*, **10**, 1100–1120.

- [3] Balakrishnan, N., Bordes, L. and Zhao, X. (2010). Minimum-distance parametric estimation under progressive Type-I censoring, *IEEE Tr. Reliab.*, **59**, 413–425.
- [4] Basu, S., Basu, A.P. and Mukhopadhyay, C. (1999). Bayesian analysis for masked system failure data using non-identical Weibull models, *J. Statist. Plann. Inference*, **78**, 255–275
- [5] Craiu, R.V. and Duchesne, T. (2004). Inference based on the EM algorithm for the competing risks model with masked causes of failure, *Biometrika*, **91**, 543–558.
- [6] Craiu, R.V. and Reiser, B. (2007). Inference for the dependent competing risks model with masked causes of failure, *Lifetime Data Analysis*, to appear.
- [7] Crowder, M. (2001). *Classical Competing Risks*, Chapman & Hall/CRC Press, London.
- [8] Dauxois, J.Y. (2000). A new method for proving weak convergence results applied to nonparametric estimators in survival analysis, *Stochastic Processes and Their Applications*, **90** (2), 327–334.
- [9] Dinse, G.E. (1986). Nonparametric prevalence and mortality estimators for animal experiments with incomplete cause-of-death data, *J. Amer. Statist. Assoc.*, **81**, 328–336.
- [10] Flehinger, B.J., Reiser, R. and Yashchin, E. (2002). Parametric modeling for survival with competing risks and masked failure causes, *Lifetime Data Analysis*, **8**, 177–203.
- [11] Fleming, T.R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis*, Wiley, New York.
- [12] Gijbels, I., Lin, D.Y. and Ying, Z. (1993). Non- and semi-parametric analysis of failure time data with missing failure indicators, *Preprint*.

- [13] Gill, R.D. and Johansen, S (1990). A survey of product-integration with a view toward application in survival analysis. *Ann. Statist.* **18** (4), 1501–1555.
- [14] Goetghebeur, E. and Ryan, L. (1995). Analysis of competing risks survival data when some failure types are missing, *Biometrika*, **82**, 821–833.
- [15] Guttman, I., Lin, D.K.J., Reiser, B. and Usher, J.S. (1995). Dependent masking and system life data analysis: Bayesian inference for two-component systems, *Lifetime Data Analysis*, **1**, 87–100.
- [16] Lin, D.K.J., Usher, J.S., Guess, F.M. (1993). Exact maximum likelihood estimation under masked system data, *IEEE Trans. Reliability*, **42**, 631–635.
- [17] Lin, D.Y. and Ying, Z. (1994). Semiparametric estimation of the additive risk model, *Biometrika*, **81**, 61–71.
- [18] Lo, S.-H. (1991). Estimating a survival function with incomplete cause-of-death data, *J. Multivariate Anal.*, **39**, 217–235.
- [19] McKeague, I.W. and Subramanian, S. (1998). Product-limit estimators and Cox regression with missing censoring information, *Scandinavian Journal of Statistics*, **25**, 589–601.
- [20] Miyakawa, M. (1984). Analysis of incomplete data in competing risks model, *IEEE Trans. Reliability*, **33**, 293–296.
- [21] Rebolledo, R. (1980). Central limit theorems for local martingales, *Z. Wahrsch. verw. Geb.*, **51**, 269–286.
- [22] Reiser, B., Flehinger, B.J. and Conn, A.R. (1996). Estimating component-defect probability from masked system success/failure data, *IEEE Trans. Reliability*, **45**, 238–243.
- [23] Schabe, H. (1994). Nonparametric estimation of component lifetime based on masked system life test data, *J. Royal Statist. Soc. Ser. B* **56**, 251–259.

- [24] Subramanian, S. (2004). Asymptotically efficient estimation of a survival function in the missing censoring indicator model. *J. Nonparametr. Stat.*, **16**, 797–817.
- [25] Usher, J.S., Hodgson, T.J., (1988). Maximum likelihood analysis of component reliability using masked system life-test data, *IEEE Trans. Reliability*, **37**, 550–555.
- [26] van der Laan, M. and McKeague, I.W. (1998). Efficient Estimation from Right-Censored Data when Failure Indicators are Missing at Random, *Ann. Statist.*, **26**, 164–182.
- [27] van de Vaart, A.W. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer, New-York.
- [28] Zhou, X. and Sun, L. (2003). Additive hazards regression with missing censoring information, *Statistica Sinica*, **13**, 1237–1257.