Ann Inst Stat Math (2011) 63:305–329
DOI 10.1007/s10463-009-0229-1

Analysis of a semiparametric mixture model for competing risks

Angelica Hernandez-Quintero · Jean-François Dupuy · Gabriel Escarela

Received: 24 June 2008 / Revised: 4 November 2008 / Published online: 25 April 2009

Abstract Semiparametric mixture regression models have recently been proposed to model competing risks data in survival analysis. In particular, Ng and McLachlan (Stat Med 22:1097–1111, 2003) and Escarela and Bowater (Commun Stat Theory Methods 37:277–293, 2008) have investigated the computational issues associated with the nonparametric maximum likelihood estimation method in a multinomial logistic/proportional hazards mixture model. In this work, we rigorously establish the existence, consistency, and asymptotic normality of the resulting nonparametric maximum likelihood estimators. We also propose consistent variance estimators for both the finite and infinite dimensional parameters in this model.

Keywords Censored failure time data · Competing risks · Large-sample properties · Maximum likelihood estimation · Mixture model · Multinomial logistic · Proportional hazards model

1 Introduction

The analysis of competing risks data is one of the most prominent areas of research in survival analysis. Such data are obtained when the failure of an individual or item

A. Hernandez-Quintero · G. Escarela
Departamento de Matemáticas, Universidad Autónoma Metropolitana, Unidad Iztapalapa, Av. San Rafael Atlixco No. 186, Col. Vicentina, 09340 Mexico, DF, Mexico
e-mail: angelicahernandezquintero@gmail.com

G. Escarela
e-mail: ge@xanum.uam.mx

A. Hernandez-Quintero · J.-F. Dupuy (✉)
Institut de Mathématiques de Toulouse, UMR 5219, Université Toulouse 3, Toulouse, France
e-mail: dupuy@math.univ-toulouse.fr

results from one of J mutually exclusive competing causes. A variety of disciplines, which include epidemiology, finance, criminology, and engineering, have given rise to a rich competing risks literature. See, for example, Klein and Bajorunaite (2004) and Klein and Moeschberger (1997) and references therein for a recent discussion of several inference problems in competing risks.

Larson and Dinse (1985) presented a precise specification of a competing risks model, which provides the basis for a concise summary of the data and enhances the understanding of the cause-specific failure-time process. This model mixes the conditional distribution of the time to failure given the failure cause over the marginal distribution of the failure cause. More precisely, let T^0 denote a random failure time and $H \in \{1, \ldots, J\}$ denote the cause of failure. For example, in a recent analysis of a prostate cancer dataset (Escarela and Bowater 2008), the failure time T^0 corresponds either to the time to death from prostate cancer, or to death from cardiovascular disease, or to death from other causes, whichever occurs first. The goal of the study was to compare the levels of a drug used to treat prostate cancer, with respect to the survival of the patients. Due to potentially fatal side effects associated with the drug, it was important, in assessing the benefit of the drug, to take into account not only the time to death from prostate cancer, but also from other competing causes. The mixture model proposed by Larson and Dinse (1985) combines parametric models for the distribution of H and the conditional distribution of T^0 given $H = j (j \in \{1, \ldots, J\})$. Larson and Dinse (1985) developed a maximum likelihood estimation procedure for their model, and Choi and Zhou (2002) and Maller and Zhou (2002) investigated large-sample properties of the resulting estimators, which include existence, consistency, and asymptotic normality.

Semiparametric generalizations of Larson and Dinse's model have recently been proposed and investigated. For example, Ng and McLachlan (2003) and Escarela and Bowater (2008) consider a semiparametric mixture model with covariates, where the marginal distribution of H conditional on the covariates has a multinomial logistic form, and the conditional distribution of T^0, given the covariates and the failure cause is specified through a proportional hazards model (Cox 1972). Ng and McLachlan (2003) and Escarela and Bowater (2008) propose and implement EM-type algorithms in this class of models. Naskar et al. (2005) consider a similar model in the case of clustered failure time data, and develop a Monte Carlo EM algorithm.

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

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

The paper is organized as follows. Section 2 describes the semiparametric mixture model for competing risks data and states some useful assumptions. In Sect. 3, we describe the maximum likelihood approach in this model and rigorously establish the existence of the resulting estimators. Section 4 and 5 establish, respectively, the consistency and asymptotic normality of these estimators. Section 5 provides consistent variance estimators for both the finite and infinite dimensional parameters in the model. We give some concluding remarks in Sect. 6. The Appendix contains the proofs of some technical lemmas.

2 Notation and model assumptions

In this section, we describe the data and the semiparametric mixture regression model derived from Larson and Dinse (1985). We also state some notations and model assumptions that will be used throughout the paper.

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

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

The semiparametric mixture model for competing risks data with covariates first relates the distribution of H to the covariate X by a multinomial logistic regression model:

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

where γ_j ($j \in \mathcal{J}$) is an unknown q-dimensional vector of parameters and γ_J is set equal to 0 for identifiability purposes. Proportional hazards models (Cox 1972) are then assumed to specify the conditional hazard of failure given H and the covariate Z; that is,

$$\lim_{h\downarrow 0}\frac{1}{h}\mathbb{P}(t < T^0 \leq t+h|T^0 > t, H=j, Z) = \lambda_j(t)\exp(\beta_j' Z), \quad j \in \mathcal{J} \tag{2}$$

where β_j ($j \in \mathcal{J}$) is a p-vector of unknown regression parameters and $\{\lambda_j(t) : t \geq 0, j \in \mathcal{J}\}$ are J unknown baseline hazard functions. The statistical problem is that of estimating the parameters β_j, γ_j, and the cumulative baseline hazard functions $\Lambda_j = \int \lambda_j$ from the incomplete data vectors $\mathbf{O}_i$, $i = 1, \ldots, n$. In practice, the coefficients β_j ($j \in \mathcal{J}$) are often the parameters of interest in the mixture model (1) and (2).

The semiparametric mixture model for competing risks has been employed by various authors. For example, Kuk (1992) analysed a heart transplant dataset (using a simplified version of model (1) and (2), where $X = Z$), whilst Ng and McLachlan (2003) and Escarela and Bowater (2008) fitted the present model to a prostate cancer dataset.

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

Note that the model (1) and (2) is related to semiparametric mixture models for survival data with cure fraction (see, among others, Kuk and Chen 1992; Taylor 1995; Sy and Taylor 2000; Peng 2003, who investigated the computational issues raised by the estimation in this class of models. See also Fang et al. (2005) and Lu (2008) who studied the large-sample properties of maximum likelihood estimators in the proportional hazards cure model).

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

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

- (C1) Conditional on Z, H, and X, the censoring time C is independent of the failure time T^0. Conditional on Z and X, C is independent of H.
- (C2) There exists a positive constant c_0 such that $\mathbb{P}(C \geq \tau|Z, X) > c_0$ almost surely.
- (C3) The hazard function of C given Z and X, $\lambda_C(s|Z, X)$, is uniformly bounded almost surely.

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

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

(C6) The covariate vectors Z and X are bounded, i.e., $\|X\| < c_1$ and $\|Z\| < c_1$ for some constant $0 < c_1 < \infty$ (where $\|\cdot\|$ denotes the Euclidean norm). The covariance matrices of Z and X are positive definite. Let $c_2 = \min_{\beta_j, j \in \mathcal{J}, \|Z\| < c_1} \exp(\beta_j' Z)$ and $c_3 = \max_{\beta_j, j \in \mathcal{J}, \|Z\| < c_1} \exp(\beta_j' Z)$.

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

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

(C8) With probability 1, there exists a positive constant c_5, such that for every $j \in \mathcal{J}$, $P_{\theta_0}[\Delta^j | T, Z, X] > c_5$.

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

Remark 1 Condition C1 ensures that no information about θ is lost by removing terms adhering to censoring from the likelihood. Condition C2 ensures that the follow-up is sufficiently long for identifying the cumulative baseline hazard functions $\Lambda_{j,0}$ on the interval $[0, \tau]$. Conditions C3–C8 are used for the identifiability of θ_0 and the asymptotics of the proposed estimators. Condition C7 ensures that the follow-up is sufficiently long (for every failure cause) so that we can estimate the $\Lambda_{j,0}$ on the entire interval $[0, \tau]$. Condition C8 ensures that for each failure cause, failures can happen and their cause be observed at any time and for any value of the covariates. Condition C9 ensures that no information about θ is lost by removing terms adhering to the marginal distributions of Z and X from the likelihood.

3 Nonparametric maximum likelihood estimation

In this paper, we assume that there are no tied failure times (this assumption is made for ease of presentation, but our results can be easily adapted to accommodate ties). Under models (1) and (2), and conditions C1–C9, the likelihood function for the parameter θ, from the observations $\mathbf{O}_i$ $(i = 1, \dots, n)$, is proportional to

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

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

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

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

over the space

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

If they exist, the resulting estimators will be referred to as nonparametric MLEs (NPMLEs), and will be noted by $\widehat{\theta}_n = (\widehat{\mathbf{B}}_n, \widehat{\mathbf{G}}_n, \widehat{\Lambda}_{j,n}; j \in \mathcal{J})$, where $\widehat{\mathbf{B}}_n = (\widehat{\beta}'_{1,n} \ldots \widehat{\beta}'_{J,n})'$ and $\widehat{\mathbf{G}}_n = (\widehat{\gamma}'_{1,n} \ldots \widehat{\gamma}'_{J-1,n})'$. In our setting, existence of the NPMLEs is ensured by the following result:

Proposition 1 *Under conditions* C1–C9*, the maximizer* $\widehat{\theta}_n$ *of* L_n *over* Θ_n *exists and is achieved.*

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

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

$$\left\{ \left(\mathbf{B}, \mathbf{G}, \Lambda_j\{t_k^j\}\right) : \mathbf{B} \in \mathcal{B}, \mathbf{G} \in \mathcal{G}, \Lambda_j\{t_k^j\} \in [0, \infty), k = 1, \ldots, |\mathcal{S}_n^j|, j \in \mathcal{J} \right\}, \quad (3)$$

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

$$\begin{aligned} &L_n(\mathbf{B}, \mathbf{G}, (\Lambda_j\{t_k^j\})_{j,k}) \\ &= \prod_{i=1}^n \left\{ \prod_{j\in\mathcal{J}} \left[\Lambda_j\{T_i\} e^{\beta_j' Z_i} \exp\left(-e^{\beta_j' Z_i} \sum_{k=1}^{|\mathcal{S}_n^j|} \Lambda_j\{t_k^j\} 1\{t_k^j \le T_i\}\right) p_{\mathbf{G},i}^{j,X} \right]^{\Delta_i^j} \right. \\ &\quad \left. \times \left[\sum_{j\in\mathcal{J}} \exp\left(-e^{\beta_j' Z_i} \sum_{k=1}^{|\mathcal{S}_n^j|} \Lambda_j\{t_k^j\} 1\{t_k^j \le T_i\}\right) p_{\mathbf{G},i}^{j,X} \right]^{1-\Delta_i} \right\} \end{aligned} \tag{4}$$

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

Assume first that $\Lambda_j\{t_k^j\} \le L < \infty$ for every $k = 1, \ldots, |\mathcal{S}_n^j|$ and $j \in \mathcal{J}$. L_n is a continuous function of the β_j, γ_j, and $\Lambda_j\{t_k^j\}$ on the compact set $\mathcal{B} \times \mathcal{G} \times [0, L]^{s_n}$, where $s_n = \sum_{j\in\mathcal{J}} |\mathcal{S}_n^j|$. Therefore, L_n achieves its maximum on this set. To show that a maximum exists on the set $\mathcal{B}\times\mathcal{G}\times[0,\infty)^{s_n}$, we show that there exists a finite L such that for all $(\mathbf{B}^L, \mathbf{G}^L, (\Lambda_j^L\{t_k^j\})_{j,k}) \in (\mathcal{B}\times\mathcal{G}\times[0,\infty)^{s_n})\backslash(\mathcal{B}\times\mathcal{G}\times[0,L]^{s_n})$, there exists a $(\mathbf{B}, \mathbf{G}, (\Lambda_j\{t_k^j\})_{j,k}) \in \mathcal{B} \times \mathcal{G} \times [0, L]^{s_n}$ which has a larger value of L_n. Consider a proof by contradiction. That is, suppose there does not exist such a L. Then for all $L < \infty$, there exists a $(\mathbf{B}^L, \mathbf{G}^L, (\Lambda_j^L\{t_k^j\})_{j,k}) \in (\mathcal{B}\times\mathcal{G}\times[0,\infty)^{s_n})\backslash(\mathcal{B}\times\mathcal{G}\times[0,L]^{s_n})$ such that for all $(\mathbf{B}, \mathbf{G}, (\Lambda_j\{t_k^j\})_{j,k}) \in \mathcal{B} \times \mathcal{G} \times [0, L]^{s_n}$, $L_n(\mathbf{B}, \mathbf{G}, (\Lambda_j\{t_k^j\})_{j,k}) \le L_n(\mathbf{B}^L, \mathbf{G}^L, (\Lambda_j^L\{t_k^j\})_{j,k})$. But we show that $L_n(\mathbf{B}^L, \mathbf{G}^L, (\Lambda_j^L\{t_k^j\})_{j,k})$ can be made arbitrarily small by increasing L, which is a contradiction. To see this, note that (4) is bounded from above by

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

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

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

tends to 0 as L tends to $+\infty$. Therefore, the upper bound of $L_n(\mathbf{B}^L, \mathbf{G}^L, (\Lambda_j^L\{t_k^j\})_{j,k})$ can be made as close to 0 as desired by increasing L, which yields a contradiction.

Therefore, for any fixed n, the maximum of L_n is obtained in the set $\mathcal{B} \times \mathcal{G} \times [0, L]^{s_n}$, for some $L < \infty$, and on this set, the maximizer $\widehat{\theta}_n$ is achieved. $\square$

For every n, the problem of maximizing L_n over (3) reduces to a finite dimensional one, since the total number s_n of jumps of the N_i $(i = 1, \ldots, n)$ is less than or equal to n. The expectation-maximization (EM) algorithm (Dempster et al. 1977) can be used to calculate the NPMLEs, as we briefly explain now [more details on the implementation of this algorithm can be found in Escarela and Bowater (2008)].

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

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

where

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

In the M-step of the EM-algorithm, we solve the complete-data score equation conditional on the observed data. In particular, a useful integral equation for $\widehat{\Lambda}_{j,n}$ (given in Lemma 1 below) can be obtained by (i) taking the derivative with respect to the jump sizes $\Lambda_j\{t_k^j\}$, of the conditional expectation of the complete-data log-likelihood given the observed data and the NPMLE, which is given by

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

(ii) Setting $(\partial l_{\widehat{\theta}_n}(\theta)/\partial \Lambda_j\{t_k^j\})|_{\theta=\widehat{\theta}_n} = 0$ and solving for $\Lambda_j\{t_k^j\}$, (iii) summing over $\{k \in \{1, \ldots, |\mathcal{S}_n^j|\} : t_k^j \leq t\}$. Calculation details are omitted.
Let $\mathbb{P}_n$ denote the empirical probability measure. Then the following holds:

Lemma 1 *The NPMLE $\widehat{\theta}_n$ satisfies the following equation for every $j \in \mathcal{J}$:*

$$\widehat{\Lambda}_{j,n}(t) = \int_0^t \frac{1}{H_n^j(s; \widehat{\theta}_n)} dG_n^j(s), \tag{5}$$

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

This equation will prove useful in establishing the consistency of the NPMLEs, as is described in the next section.

4 Consistency

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

Theorem 1 *Under conditions* C1–C9, $\|\widehat{\mathbf{B}}_n - \mathbf{B}_0\|, \|\widehat{\mathbf{G}}_n - \mathbf{G}_0\|$, *and* $\sup_{t\in[0,\tau]} |\widehat{\Lambda}_{j,n}(t) - \Lambda_{j,0}(t)|$, *for every* $j \in \mathcal{J}$, *converge to 0 almost surely as n tends to infinity.*

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

Lemma 2 *For every* $j \in \mathcal{J}$, $\limsup_n \widehat{\Lambda}_{j,n}(\tau) < \infty$ *almost surely.*

Lemma 3 *For every* $j \in \mathcal{J}$ *and* $t \in [0, \tau]$, *define*

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

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

Proof of Theorem 1 The proof consists of two steps: (i) to show that every subsequence of n contains a further subsequence where the NPMLE $\widehat{\theta}_n$ converges, and (ii) to show that the set of limits of all convergent subsequences of $\widehat{\theta}_n$ reduces to $\{\theta_0\}$.

Proof of (i) From the compactness of $\mathcal{B} \times \mathcal{G}$, every subsequence of $(\widehat{\mathbf{B}}_n, \widehat{\mathbf{G}}_n)$ has a further subsequence, say $(\widehat{\mathbf{B}}_{\phi(n)}, \widehat{\mathbf{G}}_{\phi(n)})$, which converges to some $(\mathbf{B}^*, \mathbf{G}^*)$ in $\mathcal{B} \times \mathcal{G}$. Let $j \in \mathcal{J}$. By Lemma 2 and Helly's theorem, we can find with probability 1 a subsequence $\widehat{\Lambda}_{j,\varphi(n)}$ of $\widehat{\Lambda}_{j,\phi(n)}$ and a nondecreasing right-continuous function Λ_j^* such that $\widehat{\Lambda}_{j,\varphi(n)}(t) \to \Lambda_j^*(t)$ for all $t \in [0,\tau]$ where Λ_j^* is continuous; $\widehat{\Lambda}_{j,\varphi(n)}$ is said to converge weakly to Λ_j^*. By extracting successive sub-subsequences, we can find a further subsequence $\xi(n)$ of $\varphi(n)$ in such a way that this weak convergence holds along $\xi(n)$ for every $j \in \mathcal{J}$. We now show that Λ_j^*, for $j \in \mathcal{J}$, is continuous on $[0,\tau]$. Note first that

$$\widehat{\Lambda}_{j,\xi(n)}(t) = \int_0^t \frac{\mathbb{P}_{\xi(n)}[h^j(s,\mathbf{O};\theta_0)]}{\mathbb{P}_{\xi(n)}[h^j(s,\mathbf{O};\widehat{\theta}_{\xi(n)})]}\, d\widetilde{\Lambda}_{j,\xi(n)}(s), \tag{6}$$

where $\widetilde{\Lambda}_{j,n}$ is defined in Lemma 3. It follows from the Glivenko–Cantelli property of $\{h^j(s,\mathbf{O};\theta) : s \in [0,\tau], \theta \in \Theta\}$ that (see the proof of Lemma 3)

$$\begin{aligned}
&\sup_{s\in[0,\tau]} \left|\mathbb{P}_{\xi(n)}[h^j(s,\mathbf{O};\theta_0)] - P_{\theta_0}[h^j(s,\mathbf{O};\theta_0)]\right| \longrightarrow 0 \quad a.s.,\\
&\sup_{s\in[0,\tau]} \left|\mathbb{P}_{\xi(n)}[h^j(s,\mathbf{O};\widehat{\theta}_{\xi(n)})] - P_{\theta_0}[h^j(s,\mathbf{O};\widehat{\theta}_{\xi(n)})]\right| \longrightarrow 0 \quad a.s.
\end{aligned} \tag{7}$$

Additionally, by using the bounded convergence theorem and the facts that $(\widehat{\mathbf{B}}_{\xi(n)}, \widehat{\mathbf{G}}_{\xi(n)})$ converges to $(\mathbf{B}^*, \mathbf{G}^*)$ and $\widehat{\Lambda}_{j,\xi(n)}$ converges weakly to Λ_j^*, we obtain that $P_{\theta_0}[h^j(s, \mathbf{O}; \widehat{\theta}_{\xi(n)})]$ converges to $P_{\theta_0}[h^j(s, \mathbf{O}; \theta^*)]$ for every $s \in [0, \tau]$, where $\theta^* = (\mathbf{B}^*, \mathbf{G}^*, \Lambda_j^*; j \in \mathcal{J})$. Moreover, under assumption C3, we can show that the derivative of $P_{\theta_0}[h^j(s, \mathbf{O}; \widehat{\theta}_{\xi(n)})]$ with respect to s is uniformly bounded; hence the sequence of functions $P_{\theta_0}[h^j(\cdot, \mathbf{O}; \widehat{\theta}_{\xi(n)})]$ is equicontinuous. By the Arzela–Ascoli theorem, there exists a subsequence of $\xi(n)$, say $\psi(n)$, such that $P_{\theta_0}[h^j(\cdot, \mathbf{O}; \widehat{\theta}_n)]$ converges uniformly to $P_{\theta_0}[h^j(\cdot, \mathbf{O}; \theta^*)]$ in $[0, \tau]$ along this subsequence; we can assume that this subsequence is the same for all $j \in \mathcal{J}$, by the same argument of extraction of subsequences as above. Using the latter result, Eq. (7), and the triangle inequality, we obtain that

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

uniformly in $t \in [0, \tau]$. By taking the limits on both sides of $\widehat{\Lambda}_{j,\psi(n)}(t)$ in (6), we obtain that

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

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

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

Proof of (ii) Consider the difference

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

By letting n tend to infinity, we obtain that

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

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

Since the right side of this inequality is the negative Kullback–Leibler information, we have that

$$\sum_{j\in\mathcal{J}}\Delta^j\left[\log\lambda_j^*(T) + \beta_j^{*\prime}Z - e^{\beta_j^{*\prime}Z}\Lambda_j^*(T) + \log p_{\mathbf{G}^*}^{j,X}\right]$$

$$+ (1-\Delta)\log\left(\sum_{j\in\mathcal{J}}\exp\left(-e^{\beta_j^{*\prime}Z}\Lambda_j^*(T)\right)p_{\mathbf{G}^*}^{j,X}\right)$$

$$= \sum_{j\in\mathcal{J}}\Delta^j\left[\log\lambda_{j,0}(T) + \beta_{j,0}'Z - e^{\beta_{j,0}'Z}\Lambda_{j,0}(T) + \log p_{\mathbf{G}_0}^{j,X}\right]$$

$$+ (1-\Delta)\log\left(\sum_{j\in\mathcal{J}}\exp\left(-e^{\beta_{j,0}'Z}\Lambda_{j,0}(T)\right)p_{\mathbf{G}_0}^{j,X}\right) \tag{8}$$

almost everywhere. The proof of step (ii) will be completed if we show that the equality (8) implies $\theta^* = \theta_0$, the so-called identifiability. For that purpose, let $\Delta^l = 1$ in (8), for each $l \in \mathcal{J}$ in turn. Note that this is allowed by assumption C8. This yields the following equation for almost all $t \in [0, \tau]$, $\|z\| < c_1$, $\|x\| < c_1$:

$$\log\frac{\lambda_l^*(t)}{\lambda_{l,0}(t)} + \left(\beta_l^* - \beta_{l,0}\right)'z - e^{\beta_l^{*\prime}z}\Lambda_l^*(t) + e^{\beta_{l,0}'z}\Lambda_{l,0}(t) + \log\frac{p_{\mathbf{G}^*}^{l,x}}{p_{\mathbf{G}_0}^{l,x}} = 0.$$

This equation is analogous to Eq. (A.1) in Dupuy et al. (2006). The rest of the proof of identifiability is thus essentially the same as the proof of Proposition 1 in Dupuy et al. (2006), and is omitted.

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

5 Asymptotic normality and variance estimation

5.1 Score and information

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

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

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

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

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

$$\left.\frac{\partial l_{\widehat{\theta}_n}(\widehat{\theta}_{n,\epsilon})}{\partial \epsilon}\right|_{\epsilon=0} = 0 \tag{9}$$

for every $\mathbf{h}$. Define $\Psi_{\mathbf{B}}(\theta) = (\Psi_{\beta_1}(\theta)' \ldots \Psi_{\beta_J}(\theta)')'$ and $\Psi_{\mathbf{G}}(\theta) = (\Psi_{\gamma_1}(\theta)' \ldots \Psi_{\gamma_{J-1}}(\theta)')'$, where, for every $j \in \mathcal{J}$,

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

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

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

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

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

Then, after some simple algebra, the score Eq. (9) can be re-expressed as $\Psi_n(\widehat{\theta}_n)(\mathbf{h}) = 0$, where $\Psi_n(\widehat{\theta}_n)(\mathbf{h})$ has the form

$$\Psi_n(\widehat{\theta}_n)(\mathbf{h}) = \mathbb{P}_n \left[\mathbf{h}'_{\mathbf{B}} \Psi_{\mathbf{B}}(\widehat{\theta}_n) + \mathbf{h}'_{\mathbf{G}} \Psi_{\mathbf{G}}(\widehat{\theta}_n) + \sum_{j \in \mathcal{J}} \Psi_{\Lambda_j}(\widehat{\theta}_n)(h_{\Lambda_j}) \right]. \tag{10}$$

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

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

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

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

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

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

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

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

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

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

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

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

where for any r-dimensional vector u, $u^{\otimes 2} = uu'$, and $W^j(s,\mathbf{O},\theta_0) = Y(s)e^{\beta'_{j,0}Z} S^j(\mathbf{O},\theta_0)$, $j \in \mathcal{J}$, $s \in [0,\tau]$.

Remark 3 Some of the terms in σ may be simplified by using the properties of the conditional expectation. For example, $P_{\theta_0}[W^j(s,\mathbf{O},\theta_0)]$ in $\sigma_{\Lambda_j}(\mathbf{h})$ simplifies to $P_{\theta_0}[Y(s)e^{\beta'_{j,0}Z}\Gamma^j]$. However, for variance estimation purposes, we will construct later an empirical version of σ by replacing θ_0 and P_{θ_0} by $\widehat{\theta}_n$ and $\mathbb{P}_n$, respectively, in $\sigma_{\mathbf{B}}$, $\sigma_{\mathbf{G}}$, and σ_{Λ_j}. Therefore, it is irrelevant to simplify, for instance, $P_{\theta_0}[W^j(s,\mathbf{O},\theta_0)]$ to $P_{\theta_0}[Y(s)e^{\beta'_{j,0}Z}\Gamma^j]$, since, for $i = 1,\ldots,n$, some Γ_i^j are missing and, thus, the empirical version of $P_{\theta_0}[Y(s)e^{\beta'_{j,0}Z}\Gamma^j]$ cannot be calculated.

The following lemmas state some useful properties of the score and information operators. Their proofs are given in the Appendix.

Lemma 4 *Let* $\mathbf{h} \in \mathcal{H}$. *Then* $P_{\theta_0}[\Psi_1(\theta_0)(\mathbf{h})] = 0$, *and by setting* $\sigma_{\mathbf{B}}$, $\sigma_{\mathbf{G}}$, *and* σ_{Λ_j} $(j \in \mathcal{J})$ *as above,*

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

Lemma 5 *The operator* σ *is one-to-one.*

Lemma 6 *The operator* σ *is continuously invertible.*

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

5.2 Asymptotic normality of the NPMLEs

We need some further notations to establish asymptotic normality. Let $\{e_1,\ldots,e_P\}$ be the canonical basis of $\mathbb{R}^P$, where e_m is the P-dimensional column vector with a 1 in the mth position and zeros elsewhere, for every $m = 1,\ldots,P$. We denote by

$(u, 0_Q, 0;\ j \in \mathcal{J})$ the collected vector $\mathbf{h}$ such that $\mathbf{h_B} = u$, $\mathbf{h_G} = 0_Q$, and h_{Λ_j} is identically equal to 0 for every $j \in \mathcal{J}$. Define the linear operator $\varpi : \mathbb{R}^P \to \mathbb{R}^P$ by $u \mapsto \varpi(u) = \widetilde{\sigma}_\mathbf{B}((u, 0_Q, 0;\ j \in \mathcal{J}))$. Here, ϖ is a version of $\widetilde{\sigma}_\mathbf{B}$ restricted to be a function of its first argument only, with the other arguments set equal to 0. Also, define the $(P \times P)$ matrix Σ by

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

Then the following holds:

Theorem 2 *Under conditions* C1–C9, $\sqrt{n}(\widehat{\mathbf{B}}_n - \mathbf{B}_0)$ *converges in distribution to a P-variate normal distribution with mean zero and efficient variance* Σ.

Proof of Theorem 2 Our proof follows the ideas based around the proof of Theorem 3 of Fang et al. (2005), but the technical details are substantially different. We highlight the parts that are different as follows.

Similar to the proof of Theorem 3 in Fang et al. (2005), we get that

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

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

$$\sqrt{n}\mathbf{h}'_\mathbf{B}\left(\widehat{\mathbf{B}}_n - \mathbf{B}_0\right) = \sqrt{n}\left(\Psi_n(\theta_0)(\widetilde{\sigma}(\breve{\mathbf{h}})) - P_{\theta_0}\left[\Psi_1(\theta_0)(\widetilde{\sigma}(\breve{\mathbf{h}}))\right]\right) + o_p(1), \quad (11)$$

where $\breve{\mathbf{h}} = (\mathbf{h_B}, 0_Q, 0;\ j \in \mathcal{J})$. By the central limit theorem and Lemma 4, $\sqrt{n}\mathbf{h}'_\mathbf{B}$ $(\widehat{\mathbf{B}}_n - \mathbf{B}_0)$ converges in distribution to a normal law with mean zero and variance $P_{\theta_0}[\Psi_1(\theta_0)(\widetilde{\sigma}(\breve{\mathbf{h}}))^2]$, for every $\mathbf{h_B} \in \mathbb{R}^P$. Now, noting that $\breve{\mathbf{h}} = \sigma(\widetilde{\sigma}(\breve{\mathbf{h}})) = (\sigma_\mathbf{B}(\widetilde{\sigma}(\breve{\mathbf{h}})), \sigma_\mathbf{G}(\widetilde{\sigma}(\breve{\mathbf{h}})), \sigma_{\Lambda_j}(\widetilde{\sigma}(\breve{\mathbf{h}}));\ j \in \mathcal{J})$, it follows by Lemma 4 that

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

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

Next, let $\breve{\mathbf{h}}$ be equal to $\breve{\mathbf{h}}_m = (e_m, 0_Q, 0;\ j \in \mathcal{J})$ in (11), for each $m = 1, \ldots, P$ in turn. This yields the following system of P equations:

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

where

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

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

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

Remark 4 In a competing risk analysis, the regression parameter $\mathbf{B}$ is usually the parameter of interest. We can, however, also state an asymptotic normality result for the NPMLEs of $\mathbf{G}$ and the Λ_j. Let $\varsigma : \mathbb{R}^Q \to \mathbb{R}^Q$ be defined by $\varsigma(u) = \widetilde{\sigma}_{\mathbf{G}}((0_P, u, 0; j \in \mathcal{J}))$, let $\{f_1, \ldots, f_Q\}$ be the canonical basis of $\mathbb{R}^Q$, and let $\Upsilon = (\varsigma(f_1)\ldots\varsigma(f_Q))$ be the $(Q \times Q)$ matrix of ς with respect to this basis. Also, let $\mathbf{h}_{j,t}$ be the collected vector $(\mathbf{h}_{\mathbf{B}}, \mathbf{h}_{\mathbf{G}}, h_{\Lambda_j}; j \in \mathcal{J})$ such that $\mathbf{h}_{\mathbf{B}} = 0_P$, $\mathbf{h}_{\mathbf{G}} = 0_Q$, $h_{\Lambda_j}(\cdot) = 1\{\cdot \leq t\}$ for some $t \in (0, \tau)$ and $j \in \mathcal{J}$, and h_{Λ_l} is identically equal to 0 for every $l \in \mathcal{J}, l \neq j$.

Theorem 3 *Under conditions C1–C9, $\sqrt{n}(\widehat{\mathbf{G}}_n - \mathbf{G}_0)$ converges in distribution to a Q-variate normal distribution with mean zero and variance matrix Υ. Moreover, for every $t \in (0, \tau)$ and $j \in \mathcal{J}$, $\sqrt{n}(\widehat{\Lambda}_{j,n}(t) - \Lambda_{j,0}(t))$ converges in distribution to a normal distribution with mean zero and variance $\sigma^2_{j,t} = \int_0^t \widetilde{\sigma}_{\Lambda_j}(\mathbf{h}_{j,t})(s)\, d\Lambda_{j,0}(s)$.*

Proof of Theorem 3 This result can be proved in a similar fashion to the proof of Theorem 2. The proof is therefore omitted.

5.3 Variance estimation

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

Define the $(P \times P)$, $(P \times Q)$, $(Q \times P)$, and $(Q \times Q)$ matrices $\mathbb{A}_n^{\mathbf{B}}$, $\mathbb{A}_n^{\mathbf{G}}$, $\mathbb{B}_n^{\mathbf{B}}$, and $\mathbb{B}_n^{\mathbf{G}}$ by

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

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

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

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

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

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

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

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

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

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

$$-\mathbb{P}_n \left[2\Psi_{\mathbf{B}}(\widehat{\theta}_n)' S^j(\mathbf{O}; \widehat{\theta}_n) e^{\widehat{\beta}'_{j,n} Z} Y(t_l^j) \right],$$

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

$$-\mathbb{P}_n \left[2\Psi_{\mathbf{G}}(\widehat{\theta}_n)' S^j(\mathbf{O}; \widehat{\theta}_n) e^{\widehat{\beta}'_{j,n} Z} Y(t_l^j) \right].$$

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

$$\begin{aligned} \mathbb{C}_{n,j}^{\Lambda_k}(l, m) = 1\{j = k\} \Big\{ & 1\{l = m\} \mathbb{P}_n \left[W^k(t_m^k, \mathbf{O}, \widehat{\theta}_n) \right] - \frac{2}{n} W^k(t_l^k, \mathbf{O}_{(k,m)}, \widehat{\theta}_n) \\ & + \mathbb{P}_n \left[\left\{ W^k(t_l^k, \mathbf{O}, \widehat{\theta}_n) \right\}^2 \widehat{\Delta\Lambda_{k,n}}(t_m^k) 1\{t_m^k \leq T\} \right] \Big\} + 1\{j < k\} \end{aligned}$$

$$\times \left\{ \mathbb{P}_n \left[2W^j(t_l^j, \mathbf{O}, \widehat{\theta}_n) W^k(t_l^j, \mathbf{O}, \widehat{\theta}_n) \widehat{\Delta\Lambda_{k,n}}(t_m^k) \right.\right.$$
$$\left.\left. \times \left\{ 1\{t_m^k \le T\} - 1\{t_m^k \le t_l^j\} \right\} \right] - \frac{2}{n} W^j(t_l^j, \mathbf{O}_{(k,m)}, \widehat{\theta}_n) \right\}$$

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

Define the partitioned matrix

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

and the matrices

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

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

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

and

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

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

Theorem 4 *Under conditions C1–C9, the variance estimators Σ_n, Υ_n, and $\sigma^2_{j,t,n} = \Phi'_{j,t,n} \Xi_n U_{j,t,n}$ converge in probability to Σ, Υ, and $\sigma^2_{j,t}$ ($t \in (0, \tau)$, $j \in \mathcal{J}$, respectively.*

Proof of Theorem 4 The proof of Theorem 4 relies on arguments that are now somewhat classical (see for example Parner 1998; Dupuy and Mesbah 2004; Fang et al. 2005). Hence we only provide an outline of the main steps of the proof.

First, we estimate σ by an empirical version $\sigma_n = (\sigma_{\mathbf{B},n}, \sigma_{\mathbf{G},n}, \sigma_{\Lambda_j,n};\ j \in \mathcal{J})$ obtained by replacing θ_0 and P_{θ_0} by $\widehat{\theta}_n$ and $\mathbb{P}_n$, respectively, in $\sigma_{\mathbf{B}}$, $\sigma_{\mathbf{G}}$, and σ_{Λ_j}. Similar to the proof of Theorem 3 in Parner (1998), we can show that σ_n converges in probability to σ uniformly over $\mathcal{H}$, and that its inverse $\widetilde{\sigma}_n = (\widetilde{\sigma}_{\mathbf{B},n}, \widetilde{\sigma}_{\mathbf{G},n}, \widetilde{\sigma}_{\Lambda_j,n};\ j \in \mathcal{J})$ is such that $\widetilde{\sigma}_n(\mathbf{h})$ converges to $\widetilde{\sigma}(\mathbf{h})$ in probability (see also Dupuy and Mesbah 2004). For every $\mathbf{h_B}$, the asymptotic variance of $\sqrt{n}\mathbf{h}'_{\mathbf{B}}(\widehat{\mathbf{B}}_n - \mathbf{B}_0)$ is $\mathbf{h}'_{\mathbf{B}}\varpi(\mathbf{h_B})$, which is consistently estimated by $\mathbf{h}'_{\mathbf{B}}\widetilde{\sigma}_{\mathbf{B},n}(\breve{\mathbf{h}})$, where $\breve{\mathbf{h}} = (\mathbf{h_B}, 0_Q, 0;\ j \in \mathcal{J})$. Let $\breve{\mathbf{h}}_n = (\breve{\mathbf{h}}_{\mathbf{B},n}, \breve{\mathbf{h}}_{\mathbf{G},n}, \breve{h}_{\Lambda_j,n};\ j \in \mathcal{J}) = \widetilde{\sigma}_n(\breve{\mathbf{h}})$. Then $\sigma_n(\breve{\mathbf{h}}_n) = \breve{\mathbf{h}}$, which we can write as

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

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

$$\mathbb{D}_n \begin{pmatrix} \breve{\mathbf{h}}_{\mathbf{B},n} \\ \breve{\mathbf{h}}_{\mathbf{G},n} \\ \breve{\mathbf{h}}_{\Lambda,n} \end{pmatrix} = \begin{pmatrix} \mathbf{h_B} \\ 0_Q \\ 0_{s_n} \end{pmatrix} \tag{12}$$

where $\breve{\mathbf{h}}_{\Lambda,n} = (\breve{h}_{\Lambda_1,n}(t_1^1)\ldots\breve{h}_{\Lambda_1,n}(t^1_{|\mathcal{S}_n^1|})\ldots\breve{h}_{\Lambda_J,n}(t_1^J)\ldots\breve{h}_{\Lambda_J,n}(t^J_{|\mathcal{S}_n^J|}))'$. Some algebra on (12) shows that $\breve{\mathbf{h}}_{\mathbf{B},n} = \Sigma_n\mathbf{h_B}$, with Σ_n as given above and therefore, $\mathbf{h}'_{\mathbf{B}}\Sigma_n\mathbf{h_B}$ is a consistent estimator of the asymptotic variance of $\sqrt{n}\mathbf{h}'_{\mathbf{B}}(\widehat{\mathbf{B}}_n - \mathbf{B}_0)$ for every $\mathbf{h_B}$. It follows that Σ_n is a consistent estimator of Σ. The consistency of Υ_n proceeds similarly, it is thus omitted.

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

$$\begin{cases} \sigma_{\mathbf{B},n}(\mathbf{h}_n) = 0_P \\ \sigma_{\mathbf{G},n}(\mathbf{h}_n) = 0_Q \\ \sigma_{\Lambda_j,n}(\mathbf{h}_n)(s) = 1\{s \le t\}, \quad s \in [0, \tau] \\ \sigma_{\Lambda_l,n}(\mathbf{h}_n)(s) = 0, \quad l \in \mathcal{J}, l \neq j, s \in [0, \tau]. \end{cases} \tag{13}$$

In particular, letting $s = t_1^j, \ldots, t_{|\mathcal{S}_n^j|}^j$ for every $j \in \mathcal{J}$ in (13) yields the system

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

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

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

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

6 Concluding remarks

In this paper, we have derived large-sample properties of the nonparametric maximum likelihood estimators in a semiparametric mixture regression model for competing risks data. This model was previously investigated from the computational point of view (Ng and McLachlan 2003; Escarela and Bowater 2008). Relying on modern empirical process theory, we have established the consistency and asymptotic normality of the NPMLEs. The NPML estimator of the regression parameter of interest in this model has been shown to be semiparametric efficient. Our results are asymptotic, that is, are concerned with the properties of the NPMLEs when the sample size tends to infinity. One may wonder whether these estimators are appropriate for analysing finite-sample real life data. Ng and McLachlan (2003) and Escarela and Bowater (2008) analysed a prostate cancer dataset consisting of 483 individuals. In both cases, the resulting estimates yielded coherent conclusions, which supports the conclusion that the NPMLEs are applicable for analysing real life datasets with moderate sample size.

Note that the covariate Z in model (2) was assumed to be time independent, as it was in Ng and McLachlan (2003) and Escarela and Bowater (2008). This assumption can be relaxed to accommodate time-varying covariates, provided that appropriate regularity conditions are established.

Semiparametric mixture models with applications to survival data have recently gained in popularity, and an abundant literature has been devoted to this class of models. Several issues in relation to these models are still open, however. First, it would be desirable to extend the conditional failure time model (2) to a more flexible class of models, such as the linear transformation models (see Slud and Vonta 2004 for example). It may also be interesting to accommodate more complex study designs in the statistical inference for semiparametric mixture models in competing risks data;

such designs include interval censoring and clustered failure time data. Lastly, goodness-of-fit tests in this class of models should constitute an important direction for future work.

Appendix. Proofs of technical results

Proof of Lemma 2 Let $j \in \mathcal{J}$ and $s \in [0, \tau)$. By assumption C6,

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

thus it follows from the law of large numbers that $H_n^j(s; \widehat{\theta}_n) \geq c_2 P_{\theta_0}[Y(s)\Delta^j] + o(1)$ almost surely. Under the assumptions stated in Sect. 2, $P_{\theta_0}[Y(s)\Delta^j]$ is bounded away from 0. Thus, for every $s \in [0, \tau)$, $H_n^j(s; \widehat{\theta}_n)$ is bounded away from 0 almost surely as n tends to infinity. Moreover, it is easily shown that the jump size of $\widehat{\Lambda}_{j,n}$ at τ is bounded by $1/c_2$. Therefore,

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

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

Proof of Lemma 3 We first show that the class of functions $\{h^j(s, \mathbf{O}; \theta) : s \in [0, \tau], \theta \in \Theta\}$ is Donsker. Recall that $\Theta = \mathcal{B} \times \mathcal{G} \times \mathcal{L}^{\otimes J}$. In the course of this proof, it will be useful to denote by $\mathcal{B}_j$ and $\mathcal{G}_j$ the parameter space for β_j and γ_j respectively. Consider the class

$$\mathcal{F} = \left\{ w^j(\mathbf{O}; \theta) = \frac{\exp\left(-\Lambda_j(T)e^{\beta'_j Z} + \gamma'_j X\right)}{\sum_{k \in \mathcal{J}} \exp\left(-\Lambda_k(T)e^{\beta'_k Z} + \gamma'_k X\right)} : \theta \in \Theta \right\}. \tag{14}$$

Boundedness of Z and X and Theorem 2.7.1 of van der Vaart and Wellner (1996) imply that the classes $\{g_{\beta_j}(Z) = \beta'_j Z : \beta_j \in \mathcal{B}_j\}$ and $\{g_{\gamma_j}(X) = \gamma'_j X : \gamma_j \in \mathcal{G}_j\}$ are Donsker. Differentiability of $e^{\beta'_j Z}$ in Z and boundedness of the derivative imply that $\{e^{\beta'_j Z} : \beta_j \in \mathcal{B}_j\}$ is Donsker. Moreover, the class of functions mapping T to $\Lambda_j(T)$ indexed by $\Lambda_j \in \mathcal{L}$ is also Donsker (Example 2.10.4, van der Vaart and Wellner 1996). It follows from Example 2.10.7 (van der Vaart and Wellner 1996) that the class of functions $-\Lambda_j(T)e^{\beta'_j Z} + \gamma'_j X$ with θ varying over Θ is Donsker, for every $j \in \mathcal{J}$. Then, by applying Theorem 2.10.6 of van der Vaart and Wellner (1996), we conclude that both the numerator and denominator in (14) with θ varying over Θ are Donsker classes. Since the denominator is bounded away from 0, Exemples 2.10.9 and 2.10.8 (van der Vaart and Wellner 1996) imply that $\mathcal{F}$ is Donsker. By Exemple 2.10.10 of van der Vaart and Wellner (1996), $\Delta^j + (1 - \Delta)w^j(\mathbf{O}; \theta)$ is Donsker as θ ranges over Θ. Finally, $\{Y(s) : s \in [0, \tau]\}$ is Donsker thus, by multiplying Donsker classes, we can conclude that the class $\{h^j(s, \mathbf{O}; \theta) : s \in [0, \tau], \theta \in \Theta\}$ is Donsker.

Similar arguments yield that $\{\Delta^j 1\{T \leq t\}/P_{\theta_0}[h^j(s, \mathbf{O}; \theta_0)]|_{s=T} : t \in [0, \tau]\}$ is also a Donsker class.

Next, for every $j \in \mathcal{J}$ and $t \in [0, \tau]$, define $\Lambda_j(t; \theta_0) = \int_0^t \frac{1}{P_{\theta_0}[h^j(s,\mathbf{O};\theta_0)]} P_{\theta_0} dN^j(s)$. Then

$$
\begin{aligned}
&\sup_{t\in[0,\tau]} \left|\widetilde{\Lambda}_{j,n}(t) - \Lambda_j(t; \theta_0)\right| \\
&= \sup_{t\in[0,\tau]} \left| \frac{1}{n} \sum_{i=1}^{n} \frac{\Delta_i^j 1\{T_i \leq t\}}{H_n^j(T_i; \theta_0)} - P_{\theta_0} \left[\frac{\Delta^j 1\{T \leq t\}}{P_{\theta_0}\left[h^j(s, \mathbf{O}; \theta_0)\right]|_{s=T}} \right] \right| \\
&\leq \sup_{t\in[0,\tau]} \left| \frac{1}{n} \sum_{i=1}^{n} \Delta_i^j 1\{T_i \leq t\} \left\{ \frac{1}{H_n^j(s; \theta_0)} - \frac{1}{P_{\theta_0}\left[h^j(s, \mathbf{O}; \theta_0)\right]} \right\}_{|s=T_i} \right| \\
&\quad + \sup_{t\in[0,\tau]} \left| (\mathbb{P}_n - P_{\theta_0}) \left[\frac{\Delta^j 1\{T \leq t\}}{P_{\theta_0}\left[h^j(s, \mathbf{O}; \theta_0)\right]|_{s=T}} \right] \right| \\
&\leq \sup_{s\in[0,\tau]} \left| \frac{1}{H_n^j(s; \theta_0)} - \frac{1}{P_{\theta_0}\left[h^j(s, \mathbf{O}; \theta_0)\right]} \right| \\
&\quad + \sup_{t\in[0,\tau]} \left| (\mathbb{P}_n - P_{\theta_0}) \left[\frac{\Delta^j 1\{T \leq t\}}{P_{\theta_0}\left[h^j(s, \mathbf{O}; \theta_0)\right]|_{s=T}} \right] \right|.
\end{aligned} \tag{15}
$$

From the result above, $\{h^j(s, \mathbf{O}; \theta_0) : s \in [0, \tau]\}$ is a Donsker and therefore a Glivenko–Cantelli class of functions, and thus $\sup_{s\in[0,\tau]} |H_n^j(s; \theta_0) - P_{\theta_0}[h^j(s, \mathbf{O}; \theta_0)]|$ converges to 0 almost surely. Moreover, for every $s \in [0, \tau]$, $P_{\theta_0}[h^j(s, \mathbf{O}; \theta_0)] \geq c_2 P_{\theta_0}[Y(s)\Gamma^j]$ (by C6) and thus by assumption C7, $P_{\theta_0}[h^j(s, \mathbf{O}; \theta_0)] > 0$ on $[0, \tau]$. Therefore, the first term on the right-hand side of (15) converges to 0 almost surely. The second term on the right-hand side of (15) converges almost surely to 0 by the Glivenko-Cantelli property of $\{\Delta^j 1\{T \leq t\}/P_{\theta_0}[h^j(s, \mathbf{O}; \theta_0)]|_{s=T} : t \in [0, \tau]\}$. Therefore, we conclude that $\widetilde{\Lambda}_{j,n}$ converges uniformly to $\Lambda_j(\cdot; \theta_0)$, almost surely. It is easy to verify that $\Lambda_j(\cdot; \theta_0)$ is equal to $\Lambda_{j,0}$, which concludes the proof.

Proof of Lemma 4 Let $j \in \mathcal{J}$. Then

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

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

$$\Psi_1(\theta_0)(\mathbf{h})^2 = \left[\mathbf{h}_{\mathbf{B}}'\Psi_{\mathbf{B}}(\theta_0) + \mathbf{h}_{\mathbf{G}}'\Psi_{\mathbf{G}}(\theta_0) + \sum_{j\in\mathcal{J}} \Psi_{\Lambda_j}(\theta_0)(h_{\Lambda_j})\right]^2,$$

and we take the expectation of the resulting expression. Some lengthy algebraic manipulations and re-arrangement of terms yield the result. Calculation details are omitted. □

Proof of Lemma 5 Assume $\sigma(\mathbf{h}) = 0$. By Lemma 4, $P_{\theta_0}\left[\Psi_1(\theta_0)(\mathbf{h})^2\right] = 0$, and therefore $\Psi_1(\theta_0)(\mathbf{h}) = 0$ almost surely.

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

$$h_{\Lambda_j}(t)+h_{\beta_j}'z+h_{\gamma_j}'x-\sum_{l=1}^{J-1} h_{\gamma_l}'xp_{\mathbf{G}_0}^{l,x}-e^{\beta_{j,0}'z}\left[\int_0^t h_{\Lambda_j}(s)\,d\Lambda_{j,0}(s)+h_{\beta_j}'z\Lambda_{j,0}(t)\right]=0, \tag{16}$$

with the convention $h_{\gamma_J} = 0$. By choosing t arbitrarily close to 0, (16) reduces to

$$h_{\Lambda_j}(0) + h_{\beta_j}'z + h_{\gamma_j}'x - \sum_{l=1}^{J-1} h_{\gamma_l}'xp_{\mathbf{G}_0}^{l,x} = 0, \tag{17}$$

since h_{Λ_j} and $\Lambda_{j,0}$ are continuous from the right at 0 and $\Lambda_{j,0}(0) = 0$ (by C5). Taking the difference (16) and (17) yields the following equation for almost all $t \in [0, \tau]$, $\|z\| < c_1$, and $\|x\| < c_1$:

$$h_{\Lambda_j}(t) - h_{\Lambda_j}(0) - e^{\beta_{j,0}'z}\left[\int_0^t h_{\Lambda_j}(s)\,d\Lambda_{j,0}(s) + h_{\beta_j}'z\Lambda_{j,0}(t)\right] = 0 \tag{18}$$

Let $t > 0$. Then $\Lambda_{j,0}(t) > 0$ (by C5) and Eq. (18) can be rewritten as

$$\frac{h_{\Lambda_j}(t) - h_{\Lambda_j}(0)}{\Lambda_{j,0}(t)} = e^{\beta_{j,0}'z}[r_j(t) + h_{\beta_j}'z], \tag{19}$$

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

Consider first the case where $\beta_{j,0} = 0$. Since the left-hand side of (19) does not depend on z, h_{β_j} must equal 0. Next, consider the case where $\beta_{j,0} \neq 0$. Let $t_1, t_2 > 0$. Then $e^{\beta'_{j,0}z}[r_j(t_1) - r_j(t_2)]$ should not depend on z. By assumption C6, the covariance matrix of Z is positive definite, hence we can find two distinct values z_1 and z_2 of Z such that

$$e^{\beta'_{j,0}z_1}[r_j(t_1) - r_j(t_2)] = e^{\beta'_{j,0}z_2}[r_j(t_1) - r_j(t_2)].$$

This implies that $r_j(t_1) = r_j(t_2)$, from which we deduce that $h_{\Lambda_j}(t)$ has to be constant (say, equal to c_6) for almost every $t \in (0, \tau]$. From (19), we then deduce that $h_{\Lambda_j}(0) = c_6$, which further implies that $h_{\beta_j} = 0$, $c_6 = 0$, and thus $h_{\Lambda_j}(t) = 0$ for almost every $t \in [0, \tau]$. This, together with (17) implies that $h_{\gamma_j} = 0$ for every $j = 1, \ldots, J-1$.

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

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

Proof of Lemma 6 Since $\mathcal{H}$ is a Banach space, to prove that σ is continuously invertible, it is sufficient to prove that σ is one-to-one and that it can be written as the sum $A + (\sigma - A)$ of a bounded linear operator A with bounded inverse and a compact operator $\sigma - A$ [Lemma 25.93 of van der Vaart (1998)].

σ is one-to-one by Lemma 5. Next, define the linear operator $A : \mathcal{H} \to \mathcal{H}$ by $A(\mathbf{h}) = (\mathbf{h_B}, \mathbf{h_G}, h_{\Lambda_j}(\cdot)P_{\theta_0}[W^j(\cdot, \mathbf{O}, \theta_0)];\ j \in \mathcal{J})$. A is bounded (by C4 and C6). In addition, for every $j \in \mathcal{J}$, $P_{\theta_0}[W^j(\cdot, \mathbf{O}, \theta_0)]$ is uniformly bounded away from 0 on $[0, \tau]$ (by C2, C5, C6). Thus A is invertible with bounded inverse $A^{-1}(\mathbf{h}) = (\mathbf{h_B}, \mathbf{h_G}, h_{\Lambda_j}(\cdot)P_{\theta_0}[W^j(\cdot, \mathbf{O}, \theta_0)]^{-1};\ j \in \mathcal{J})$.

Moreover, it is not difficult to show that the operator $\sigma - A$ is compact by using the same techniques as in Lu (2008) for example (we omit the details). This completes the proof.

Acknowledgments The authors gratefully acknowledge the helpful comments on an earlier version from the Associate Editor and a Referee. This research was supported by the ECOS-Nord (Evaluation-orientation de la COopération Scientifique) program M06M01 and by CONACYT (for A. Hernández-Quintero and G. Escarela). Part of this research was conducted while A. Hernández-Quintero was visiting the Institut de Mathématiques de Toulouse, with ECOS-Nord support.

References

Chang, I.-S., Hsuing, C. A., Wang, M.-C., Wen, C.-C. (2005). An asymptotic theory for the nonparametric maximum likelihood estimator in the Cox gene model. *Bernoulli, 11*, 863–892.

Choi, K. C., Zhou, X. (2002). Large sample properties of mixture models with covariates for competing risks. *Journal of Multivariate Analysis, 82*, 331–366.

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

Dempster, A. P., Laird, N. M., Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39*, 1–38.

Dupuy, J.-F., Escarela, G. (2007). Modélisation de risques concurrents par un modèle de mélange semi-paramétrique. *Comptes Rendus Mathématique. Académie des Sciences, Paris, 344*, 641–644.

Dupuy, J.-F., Mesbah, M. (2004). Estimation of the asymptotic variance of semiparametric maximum likelihood estimators in the Cox model with a missing time-dependent covariate. *Communications in Statistics: Theory and Methods, 33*, 1385–1401.

Dupuy, J.-F., Grama, I., Mesbah, M. (2006). Asymptotic theory for the Cox model with missing time-dependent covariate. *The Annals of Statistics, 34*, 903–924.

Escarela, G., Bowater, R. (2008). Fitting a semi-parametric mixture model for competing risks in survival data. *Communications in Statistics: Theory and Methods, 37*, 277–293.

Fang, H.-B., Li, G., Sun, J. (2005). Maximum likelihood estimation in a semiparametric logistic/proportional-hazards mixture model. *Scandinavian Journal of Statistics, 32*, 59–75.

Klein, J. P., Bajorunaite, R. (2004). Inference for competing risks. In *Advances in survival analysis, Handbook of statistics* (Vol. 23, pp. 291–311). Amsterdam: Elsevier.

Klein, J. P., Moeschberger, M. L. (1997). *Survival analysis: methods for censored and truncated data*. New York: Springer.

Kosorok, M. R., Song, R. (2007). Inference under right censoring for transformation models with a change-point based on a covariate threshold. *The Annals of Statistics, 35*, 957–989.

Kuk, A. Y. C. (1992). A semiparametric mixture model for the analysis of competing risks data. *Australian & New Zealand Journal of Statistics, 34*, 169–180.

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

Larson, M. G., Dinse, G. E. (1985). A mixture model for the regression analysis of competing risks data. *Journal of the Royal Statistical Society, Series C, 34*, 201–211.

Lu, W. (2008). Maximum likelihood estimation in the proportional hazards cure model. *Annals of the Institute of Statistical Mathematics, 60*, 545–574.

Lu, W., Peng, L. (2008). Semiparametric analysis of mixture regression models with competing risks data. *Lifetime Data Analysis, 14*, 231–252.

Maller, R. A., Zhou, X. (2002). Analysis of parametric models for competing risks. *Statistica Sinica, 12*, 725–750.

Murphy, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *The Annals of Statistics, 22*, 712–731.

Murphy, S. A. (1995). Asymptotic theory for the frailty model. *The Annals of Statistics, 23*, 182–198.

Naskar, M., Das, K., Ibrahim, J. G. (2005). A semiparametric mixture model for analyzing clustered competing risks data. *Biometrics, 61*, 729–737.

Ng, S. K., McLachlan, G. J. (2003). An EM-based semi-parametric mixture model approach to the regression analysis of competing-risks data. *Statistics in Medicine, 22*, 1097–1111.

Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *The Annals of Statistics, 26*, 183–214.

Peng, Y. (2003). Fitting semiparametric cure models. *Computational Statistics & Data Analysis, 41*, 481–490.

Slud, E. V., Vonta, F. (2004). Consistency of the NPML estimator in the right-censored transformation model. *Scandinavian Journal of Statistics, 31*, 21–41.

Sy, J. P., Taylor, J. M. G. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics, 56*, 227–236.

Taylor, J. M. G. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics, 51*, 899–907.

Tsiatis, A. A. (2006). *Semiparametric theory and missing data*. New York: Springer.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge: Cambridge University Press.

van der Vaart, A. W., Wellner, J. A. (1996). *Weak convergence and empirical processes*. New York: Springer.