doi: 10.1111/j.1467-9469.2007.00567.x
 Vol 34: 870–895, 2007

Non-parametric Maximum-Likelihood Estimation in a Semiparametric Mixture Model for Competing-Risks Data

I-SHOU CHANG

Institute of Cancer Research and Division of Biostatistics and Bioinformatics, National Health Research Institutes, Taiwan

CHAO A. HSIUNG

Division of Biostatistics and Bioinformatics, National Health Research Institutes, Taiwan

CHI-CHUNG WEN

Department of Mathematics, Tamkang University

YUH-JENN WU

Department of Applied Mathematics, Chung Yuan Christian University

CHE-CHI YANG

Department of Information Management, Lunghwa University of Science and Technology

ABSTRACT. This paper describes our studies on non-parametric maximum-likelihood estimators in a semiparametric mixture model for competing-risks data, in which proportional hazards models are specified for failure time models conditional on cause and a multinomial model is specified for the marginal distribution of cause conditional on covariates. We provide a verifiable identifiability condition and, based on it, establish an asymptotic profile likelihood theory for this model. We also provide efficient algorithms for the computation of the non-parametric maximum-likelihood estimate and its asymptotic variance. The success of this method is demonstrated in simulation studies and in the analysis of Taiwan severe acute respiratory syndrome data.

Key words: case fatality rate, competing-risks problems, self-consistency equation, severe acute respiratory syndrome

1. Introduction

This paper provides an asymptotic theory and efficient algorithms for the non-parametric maximum-likelihood estimate (NPMLE) in a semiparametric mixture model for competing-risks data, in which proportional hazards models are specified for failure time models conditional on cause type and a multinomial model is specified for the marginal distribution of cause type conditional on covariates.

This model appeared in Fine (1999) and is useful in the study of covariate-specific probability of failure over time from a given cause. A recent example is provided by the epidemiology of severe acute respiratory syndrome (SARS), an epidemic that affected the Asian Pacific region severely in 2003. Because of its high transmissibility and lethality, a patient suspected of having SARS is admitted to a SARS-dedicated hospital and kept in isolation immediately. There he/she may either get well and discharged after certain period of time or die with death attributed to SARS or other underlying diseases, and it is hard to know in advance which one of these would eventually happen, because of no specific cure. To study transmission dynamics and to have good planning of patient-care capacity, it is necessary to get good estimates of the case fatality rate, the distribution of the admission-to-discharge and that of the admission-to-death. In fact, assessing the case fatality rate of SARS patients has been a concern since the outbreak of the disease, and was studied by Donnelly *et al.* (2003) during the epidemic. We will fit Taiwan SARS data to the competing-risks model so that these epidemiologically important quantities can be studied directly.

Let T_i, C_i, Z_i and W_i be the time-to-event, the censoring time, the covariate and the curability of the ith individual. Let $W_i \in \{1,2\}$ with $W_i = 1$ indicating the ith individual being incurable and $W_i = 2$, curable. We assume $T_i \geq 0$, $C_i \geq 0$ and Z_i in $\Re^d$. We assume that, for $j = 1,2$ and z in $\Re^d$, the conditional hazard of T_i at t, given $W_i = j$ and $Z_i = z$, exists and is

$$\lambda_j(t)\,\mathrm{e}^{\beta_j^T z}, \tag{1}$$

where $\lambda_j(\cdot)$ is a non-negative deterministic baseline function, β_j is in $\Re^d$ and β_j^{T} is the transpose of β_j. In competing-risks problems, W_i is usually referred to as the cause of failure or failure type variable. In this paper, we may refer to T_i as the death time if $W_i = 1$ and the cure time if $W_i = 2$.

Assume (T_i, W_i) and C_i are conditionally independent, given Z_i. Denoted by $F_j(t,z)$, the conditional distribution of T_i at t, given $W_i = j$ and $Z_i = z$. Denoted by $G(\cdot, z)$, the conditional distribution of C_i, given $Z_i = z$. Let $[\triangle_i = 1] = [T_i \leq C_i, W_i = 1]$, $[\triangle_i = 2] = [T_i \leq C_i, W_i = 2]$, and $[\triangle_i = 3] = [T_i > C_i]$. Assume that, for some $\alpha_1 \in \Re^1$ and $\alpha_2 \in \Re^d$,

$$P(W_i = 1 \mid Z_i = z) = \frac{\mathrm{e}^{\alpha_1 + \alpha_2^T z}}{1 + \mathrm{e}^{\alpha_1 + \alpha_2^T z}}, \tag{2}$$

which will be denoted by $\alpha(z)$. Let $X_i = T_i \wedge C_i$. Then the likelihood for $(X_i, \Delta_i) = (x, \delta)$ given $Z_i = z$ is

$$\begin{aligned} &\{\alpha(z)(1 - G(x,z))f_1(x,z)\}^{[\delta=1]}\{(1-\alpha(z))(1-G(x,z))f_2(x,z)\}^{[\delta=2]} \\ &\times \{\alpha(z)(1-F_1(x,z))g(x,z) + (1-\alpha(z))(1-F_2(x,z))g(x,z)\}^{[\delta=3]}, \end{aligned} \tag{3}$$

where $g(x,z) \equiv \frac{\partial G}{\partial x}(x,z)$ and $f_j(x,z) \equiv \frac{\partial F_j}{\partial x}(x,z)$, which are assumed to exist.

Let $\alpha_c \equiv (\alpha_1, \alpha_2^T)^T$ and $\Lambda_j(t) = \int_0^t \lambda_j(s)\,\mathrm{d}s$. We will study the NPMLE of $\theta = (\alpha_c^T, \beta_1^T, \beta_2^T, \Lambda_1, \Lambda_2)$ based on $\{X_i, \triangle_i, Z_i \mid i = 1, \ldots, n\}$ and apply it to Taiwan SARS data, assuming that (T_i, C_i, Z_i, W_i) is an independent and identically distributed sequence for $i = 1, \ldots, n$.

We note that Fine (1999) studied the above model under the additional assumption that C_1 is independent of (T_1, W_1, Z_1) and, among other things, provided robust estimates for α_c, β_1 and β_2 by first obtaining an estimate of the distribution of C_1.

Other works for competing-risks data having an explicit cumulative incidence function like (2) include the parametric mixture model of Larson & Dinse (1985) and the non-parametric mixture model of Betensky & Schoenfeld (2001); Maller & Zhou (2002) and Choi & Zhou (2002) systematically studied the models of Larson & Dinse (1985); Betensky & Schoenfeld (2001) examined non-parametric estimation in their mixture model.

An alternative approach is provided by Fine & Gray (1999) and Fine (2001); see also Andersen (2002) for a multi-state formulation of competing risks model. They modelled directly the crude failure probability, or cumulative incidence function $P(T_1 \leq t, W_1 = 1 \mid Z_1)$ and developed estimation procedures in their semiparametric models. In particular, an inverse probability of censoring weighting technique (Robins & Rotnitzky, 1992) is employed for analyzing right-censored competing-risks data, in which a consistent estimate of the distribution G of C_1 conditional on Z_1 is needed.

As mentioned in Fine (1999), the approach with explicit specification such as (2) has precedent in the cure-rate model; see Farewell (1977). Readers are referred to Kuk (1992), Kuk & Chen (1992), Sy & Taylor (2000), Peng & Dear (2000), Peng (2003) and references therein for semiparametric cure-rate models. We would like to point out that Taylor (1995) and Li *et al.* (2001) made a thorough study of identifiability in these cure-rate models.

The results of this paper rest on an identifiability assumption, which says roughly that identifiability is guaranteed if there are enough score functions. We provide a method to verify this assumption and illustrate this method by showing that the model is identifiable if $d=1$, Z_1 takes at least three different values and there exists covariate effect.

For the consistency, we first present integral equations, derived from score functions, for the NPMLE and then apply empirical process theory to establish it, following the approaches and techniques developed in Murphy (1994), Murphy *et al.* (1997), Parner (1998), Kosorok *et al.* (2004) and Chang *et al.* (2005), among others. With the consistency, we then treat the NPMLE as an Z-estimator and prove its asymptotic normality. Finally, we provide estimates of the asymptotic variance for the NPMLE of α_c, β_1 and β_2 by the observed profile information developed in Murphy & van der Vaart (1999). In fact, we followed Murphy & van der Vaart (2000) for developing a profile likelihood theory.

The simulation studies use the algorithms based on the integral equations and the profile likelihood theory. We find that when C_1 and (T_1, W_1, Z_1) are independent, both the method in Fine (1999) and ours work nicely, although our method seems to be a little better in terms of mean-squared error; when C_1 and (T_1, W_1, Z_1) are not independent but satisfy the assumptions in this paper, our method still works but the method in Fine (1999) may fail.

This paper is organized as follows. Section 2 presents the likelihood function, discusses the identifiability assumption, establishes the existence of NPMLE, derives the score functions and the integral equations for the NPMLE, and presents the algorithm based on the integral equations. Sections 3, 4 and 5, respectively, establish the consistency, the asymptotical normality and the profile likelihood theory. Section 6 illustrates the method in simulation studies. Section 7 applies the method to analyze Taiwan SARS data. Section 8 gives concluding remarks. The Appendix contains the proofs concerning model identifiability and the more technical proofs used in establishing the consistency.

2. Identifiability and the score functions

2.1. Identifiability

The parameter space we consider is

$$\Theta=\{\theta=(\alpha_c^T,\beta_1^T,\beta_2^T,\Lambda_1,\Lambda_2)\,|\,\alpha_c\in\mathcal{A},\beta_1\in\mathcal{B},\beta_2\in\mathcal{B},\Lambda_1\in\mathcal{L},\Lambda_2\in\mathcal{L}\}.$$

Here, $\mathcal{A}$ is a compact subset of $\Re^{d+1}$ with non-empty interior, $\mathcal{B}$ is a compact subset of $\Re^d$ with non-empty interior, and

$$\mathcal{L}=\{\Lambda:[0,\tau]\rightarrow[0,\infty)\,|\,\Lambda(0)=0,\Lambda \text{ is non-decreasing and right continuous}\}$$

for some $\tau>0$. The analysis of T_i in this paper is restricted to the interval $[0,\tau]$. Denote the true parameter $(\alpha_{c0}^T,\beta_{10}^T,\beta_{20}^T,\Lambda_{10},\Lambda_{20})$ by θ_0. Here α_{c0},β_{10}, and β_{20} are assumed to be interior points of $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{B}$, respectively; and Λ_{10} and Λ_{20} have positive and bounded derivatives on $[0,\tau]$; the assumption that Λ_{10} and Λ_{20} have positive derivative on $[0,\tau]$ is made to simplify the presentation and some extension is possible. We assume $P(T_i>\tau, W_i=1\,|\,Z_i)>0$ a.s., $P(T_i>\tau, W_i=2\,|\,Z_i)>0$ a.s., and $P(C_i\geq\tau\,|\,Z_i)>0$ a.s. We further assume that the support of Z_i is bounded and non-degenerate. We note that it loses no generality and simplifies the presentation to assume $P(C_i \le \tau \mid Z_i) = 1$ a.s. Assume further that the distribution of (C_i, Z_i) has nothing to do with θ.

It follows from (3) that the likelihood for the ith individual is

$$\begin{aligned}\tilde{L}_{(i),\,\theta} \equiv & \{\alpha(Z_i)\lambda_1(X_i)\exp(\beta_1^T Z_i - \Lambda_1(X_i)\,\mathrm{e}^{\beta_1^T Z_i})\}^{[\triangle_i = 1]} \\ & \times \{[1-\alpha(Z_i)]\lambda_2(X_i)\exp(\beta_2^T Z_i - \Lambda_2(X_i)\,\mathrm{e}^{\beta_2^T Z_i})\}^{[\triangle_i = 2]} \\ & \times \{\alpha(Z_i)\exp(-\Lambda_1(X_i)\,\mathrm{e}^{\beta_1^T Z_i}) + [1-\alpha(Z_i)]\exp(-\Lambda_2(X_i)\,\mathrm{e}^{\beta_2^T Z_i})\}^{[\triangle_i = 3]},\end{aligned}$$

where λ_1 and λ_2 are, respectively, the derivatives of Λ_1 and Λ_2.

In this paper, we need the following.

Assumption 1 (Identifiability)

If Λ_1 is absolutely continuous relative to Λ_{10} and Λ_2 is absolutely continuous relative to Λ_{20}, then $\tilde{L}_{(1),\,\theta} = \tilde{L}_{(1),\,\theta_0}$ a.s. implies $\theta = \theta_0$.

The condition $\tilde{L}_{(1),\,\theta} = \tilde{L}_{(1),\,\theta_0}$ a.s. in assumption 1 implies that

$$\begin{aligned}& \{\alpha(Z_1)\,\frac{\mathrm{d}\Lambda_1}{\mathrm{d}\Lambda_{10}}(X_1)\exp(\beta_1^T Z_1 - \Lambda_1(X_1)\,\mathrm{e}^{\beta_1^T Z_1})\}^{[\triangle_1 = 1]} \\ & \quad \times \{[1-\alpha(Z_1)]\,\frac{\mathrm{d}\Lambda_2}{\mathrm{d}\Lambda_{20}}(X_1)\exp(\beta_2^T Z_1 - \Lambda_2(X_1)\,\mathrm{e}^{\beta_2^T Z_1})\}^{[\triangle_1 = 2]} \\ & \quad \times \{\alpha(Z_1)\exp(-\Lambda_1(X_1)\,\mathrm{e}^{\beta_1^T Z_1}) + [1-\alpha(Z_1)]\exp(-\Lambda_2(X_1)\,\mathrm{e}^{\beta_2^T Z_1})\}^{[\triangle_1 = 3]} \\ & = \{[\alpha_0(Z_1)]\exp(\beta_{10}^T Z_1 - \Lambda_{10}(X_1)\,\mathrm{e}^{\beta_{10}^T Z_1})\}^{[\triangle_1 = 1]} \\ & \quad \times \{[1-\alpha_0(Z_1)]\exp(\beta_{20}^T Z_1 - \Lambda_{20}(X_1)\,\mathrm{e}^{\beta_{20}^T Z_1})\}^{[\triangle_1 = 2]} \\ & \quad \times \{\alpha_0(Z_1)\exp(-\Lambda_{10}(X_1)\,\mathrm{e}^{\beta_{10}^T Z_1}) + [1-\alpha_0(Z_1)]\exp(-\Lambda_{20}(X_1)\,\mathrm{e}^{\beta_{20}^T Z_1})\}^{[\triangle_1 = 3]}, \text{ a.s.}\end{aligned} \tag{4}$$

In view of the left-hand side of (4), we define

$$\begin{aligned}& \tilde{L}_{(1),\,(\alpha_c,\,\beta_1,\,\beta_2,\,y_{10},\,y_{20},\,y_1,\,y_2)}(\Delta_1, Z_1) \\ & \quad = \left\{\alpha(Z_1) y_1 \exp(\beta_1^T Z_1 - y_{10}\mathrm{e}^{\beta_1^T Z_1})\right\}^{[\triangle_1 = 1]} \left\{[1-\alpha(Z_1)] y_2 \exp(\beta_2^T Z_1 - y_{20}\mathrm{e}^{\beta_2^T Z_1})\right\}^{[\triangle_1 = 2]} \\ & \quad\quad \left\{\alpha(Z_1)\exp(-y_{10}\mathrm{e}^{\beta_1^T Z_1}) + [1-\alpha(Z_1)]\exp(-y_{20}\mathrm{e}^{\beta_2^T Z_1})\right\}^{[\triangle_1 = 3]},\end{aligned}$$

and consider the following.

Assumption 2

There exists t^* in the support of the conditional distribution of C_1 given Z_1 such that

$$\tilde{L}_{(1),\,(\alpha_c,\,\beta_1,\,\beta_2,\,y_{10},\,y_{20},\,y_1,\,y_2)}(\Delta_1, Z_1) = \tilde{L}_{(1),\,(\alpha_{c0},\,\beta_{10},\,\beta_{20},\,\Lambda_{10}(t^*),\,\Lambda_{20}(t^*),\,1,\,1)}(\Delta_1, Z_1), \textit{ a.s.}$$

implies $\alpha_c = \alpha_{c0}, \beta_1 = \beta_{10}, \beta_2 = \beta_{20}, y_{10} = \Lambda_{10}(t^*), y_{20} = \Lambda_{20}(t^*), y_1 = 1, y_2 = 1.$

We will show, in the Appendix, that assumption 1 is implied by assumption 2, which involves only 'parameters' in finite dimensional space. The following assumption 2′ implies assumption 2 locally, by the inverse function theorem, and can be verified numerically.

Assumption 2′

There exists t^* in the support of the conditional distribution of C_1 given Z_1 such that if

$$\left(a_1^T\frac{\partial}{\partial\alpha_c}+a_2^T\frac{\partial}{\partial\beta_1}+a_3^T\frac{\partial}{\partial\beta_2}+a_4\frac{\partial}{\partial y_{10}}+a_5\frac{\partial}{\partial y_{20}}+a_6\frac{\partial}{\partial y_1}+a_7\frac{\partial}{\partial y_2}\right)\log\{\tilde{L}_{(1),(\alpha_c,\beta_1,\beta_2,y_{10},y_{20},y_1,y_2)}(\Delta_1,Z_1)\}=0$$

when evaluated at

$$(\alpha_c,\beta_1,\beta_2,y_{10},y_{20},y_1,y_2)=(\alpha_{c0},\beta_{10},\beta_{20},\Lambda_{10}(t^*),\Lambda_{20}(t^*),1,1)$$

for every possible values of $(\triangle_1,Z_1)$, then $a_1=0$ in $\Re^{d+1}$, $a_2=a_3=0$ in $\Re^d$, and $a_4=a_5=a_6=a_7=0$ in $\Re$.

We now illustrate the use of assumption 2′ by assuming $d=1$, $P(Z_1=0)>0$, $P(Z_1=0.5)>0$, and $P(Z_1=1)>0$.

Let $\mathcal{F}:\mathcal{N}\mapsto\Re^8$ be the function whose components are of the form

$$(\alpha_c,\beta_1,\beta_2,y_{10},y_{20},y_1,y_2)\mapsto\log\left\{\tilde{L}_{(1),(\alpha_c,\beta_1,\beta_2,y_{10},y_{20},y_1,y_2)}(\Delta_1,Z_1)\right\},$$

with the components defined by $(\Delta_1,Z_1)\in\{(1,0),(1,0.5),(1,1),(2,0),(2,0.5),(2,1),(3,0),(3,1)\}$. Here $\mathcal{N}$ is a neighbourhood of $(\alpha_{c0},\beta_{10},\beta_{20},\Lambda_{10}(t^*),\Lambda_{20}(t^*),1,1)$ in $\Re^8$. We can validate assumption 2′ by showing that the determinant of the Jacobian of $\mathcal{F}$, $\det(J_{\mathcal{F}})$, at $(\alpha_{c0},\beta_{10},\beta_{20},\Lambda_{10}(t^*),\Lambda_{20}(t^*),1,1)$ is not zero for some $0\le t^*\le\tau$ in the support of the conditional distribution of C_1 given Z_1. Suppose $\alpha_{c0}=(\alpha_{10},\alpha_{20})^T=(-2,5)^T$, $\beta_{10}=0.5$, $\beta_{20}=-0.5$, $\Lambda_{10}(t)=t/4$, $\Lambda_{20}(t)=t/5$, $t^*=1$, then computer calculation gives $\det(J_{\mathcal{F}})=0.0085$. This shows that assumption 2′ is satisfied for $\alpha_{c0}=(-2,5)^T,\beta_{10}=0.5$, $\beta_{20}=-0.5,\Lambda_{10}(t)=t/4,\Lambda_{20}(t)=t/5$.

In fact, we computed $\det(J_{\mathcal{F}})$ for many values of $(\alpha_{c0},\beta_{10},\beta_{20},\Lambda_{10}(t^*),\Lambda_{20}(t^*))$. Based on these calculations, we wish to conjecture that, in the case $d=1$ and Z_1 can take three different values, assumption 2′ holds if and only if at least one of $\alpha_{20},\beta_{10},\beta_{20}$ is not zero. This conjecture seems to be in line with the theorems 1 and 2 in Li *et al.* (2001).

We note that a referee pointed out that when $\alpha_{20}=\beta_{10}=\beta_{20}=0$, the model is not identifiable. The argument provided by the referee is as follows. Suppose that covariates have no effect at all and $C=\tau$. Then the likelihood function becomes

$$(\alpha_1\lambda_1(X_i)\exp\{-\Lambda_1(X_i)\})^{[\Delta_i=1]}((1-\alpha_1)\lambda_2(X_i)\exp\{-\Lambda_2(X_i)\})^{[\Delta_i=2]}$$
$$\{\alpha_1\exp\{-\Lambda_1(\tau)\}+(1-\alpha_1)\exp\{-\Lambda_2(\tau)\}\}^{[\Delta_i=3]};$$

and for any $1-\exp\{-\Lambda_1(\tau)\}<\gamma<1$, it is easy to check that

$$\tilde{\alpha}_1=\gamma\alpha_1,\tilde{\Lambda}_1=-\log\{\exp\{-\Lambda_1\}+\gamma-1\}+\log\gamma,$$
$$\tilde{\Lambda}_2=-\log\{\exp\{-\Lambda_2\}+(1-\gamma\alpha_1)/(1-\alpha_1)-1\}+\log((1-\gamma\alpha_1)/(1-\alpha_1))$$

give the same likelihood function. This shows the non-identifiability. This also indicates that the method of this paper cannot be used to test the hypothesis that there exists convariate effects.

We also computed $\det(J_{\mathcal{F}})$ when $d=2$. The results seem to suggest that the above conjecture remains valid; namely, assumption 2′ holds if and only if at least one of $\alpha_{20},\beta_{10},\beta_{20}$ is not zero as a vector in $\Re^2$. We are interested in knowing if this conjecture can be confirmed and how general it is. Since assumption 2′ is local in nature, it is important in practice to know how large Θ can be to still make assumption 1 valid.

2.2. Score functions

Because $\tilde{L}_{(i),\theta}$ could become arbitrarily large within the class of absolutely continuous Λ_1 and Λ_2, we consider the likelihood

$$L_n(\theta) \equiv \prod_{i=1}^{n} \bar{L}_{(i),\theta} \tag{5}$$

with

$$\begin{aligned}\bar{L}_{(i),\theta} = &\{\alpha(Z_i)\triangle\Lambda_1(X_i)\exp(\beta_1^T Z_i - \Lambda_1(X_i)\mathrm{e}^{\beta_1^T Z_i})\}^{[\triangle_i=1]}\\ &\times\{[1-\alpha(Z_i)]\triangle\Lambda_2(X_i)\exp(\beta_2^T Z_i - \Lambda_2(X_i)\mathrm{e}^{\beta_2^T Z_i})\}^{[\triangle_i=2]}\\ &\times\{\alpha(Z_i)\exp(-\Lambda_1(X_i)\mathrm{e}^{\beta_1^T Z_i}) + [1-\alpha(Z_i)]\exp(-\Lambda_2(X_i)\mathrm{e}^{\beta_2^T Z_i}\}^{[\triangle_i=3]}\end{aligned}$$

and $\triangle\Lambda_1(t) = \Lambda_1(t) - \Lambda_1(t-)$ and $\triangle\Lambda_2(t) = \Lambda_2(t) - \Lambda_2(t-)$.

The NPMLE $\hat{\theta}_n \equiv (\hat{\alpha}_{cn}^T, \hat{\beta}_{1n}^T, \hat{\beta}_{2n}^T, \hat{\Lambda}_{1n}, \hat{\Lambda}_{2n})$ we propose is the maximizer of (5) over $\mathcal{A}\times\mathcal{B}\times\mathcal{B}\times\mathcal{L}_*\times\mathcal{L}_*$, where $\mathcal{L}_*\subset\mathcal{L}$ comprises step functions. In fact, $\hat{\Lambda}_{1n}$ has positive jumps only at X_i with $\triangle_i = 1$ and $\hat{\Lambda}_{2n}$ has positive jumps only at X_i with $\triangle_i = 2$.

We assume all the random variables are defined on a sample space Ω with a specific σ-field. Let $\omega\in\Omega$ and n be fixed. Using the compactness of $\mathcal{A}\times\mathcal{B}\times\mathcal{B}$ and the fact that $\lim_{y\to\infty} ye^{-y} = 0$, we conclude immediately from

$$\begin{aligned}|L_n(\theta)| \le \prod_{i=1}^{n} &\{\alpha(Z_i)\Lambda_1(X_i)\exp(\beta_1^T Z_i - \Lambda_1(X_i)\mathrm{e}^{\beta_1^T Z_i})\}^{[\triangle_i=1]}\\ &\times\{[1-\alpha(Z_i)]\Lambda_2(X_i)\exp(\beta_2^T Z_i - \Lambda_2(X_i)\mathrm{e}^{\beta_2^T Z_i})\}^{[\triangle_i=2]}\\ &\times\{\alpha(Z_i)\exp(-\Lambda_1(X_i)\mathrm{e}^{\beta_1^T Z_i}) + [1-\alpha(Z_i)]\exp(-\Lambda_2(X_i)\mathrm{e}^{\beta_2^T Z_i})\}^{[\triangle_i=3]}\end{aligned}$$

that the following theorem holds.

Theorem 1
The NPMLE $\hat{\theta}_n$ exists, and for every $n = 1, 2, \ldots,$ there is a constant M_n such that $\hat{\Lambda}_{1n}(\tau) < M_n$ and $\hat{\Lambda}_{2n}(\tau) < M_n$.

Let $BV[0,\tau]$ denote the set of all real-valued functions on $[0,\tau]$ with finite variation. For $h_1\in\Re^{d+1}$, $h_2, h_3\in\Re^d$, and $h_4, h_5\in BV[0,\tau]$, denote by $\ell_{1,\theta}[h_1]$, $\ell_{2,\theta}[h_2]$, $\ell_{3,\theta}[h_3]$, $\ell_{4,\theta}[h_4]$, and $\ell_{5,\theta}[h_5]$, respectively, the score functions for the submodels specified by $\alpha_c + \varepsilon h_1$, $\beta_1 + \varepsilon h_2$, $\beta_2 + \varepsilon h_3$, $\Lambda_{1\varepsilon}(\cdot) = \int_0^{\cdot}(1+\varepsilon h_4)\,\mathrm{d}\Lambda_1$ and $\Lambda_{2\varepsilon}(\cdot) = \int_0^{\cdot}(1+\varepsilon h_5)\,\mathrm{d}\Lambda_2$, with ε near 0. Let

$$V_{1i}(\theta) = \frac{\alpha(Z_i)\exp(-\mathrm{e}^{\beta_1^T Z_i}\Lambda_1(X_i))}{\alpha(Z_i)\exp(-\mathrm{e}^{\beta_1^T Z_i}\Lambda_1(X_i)) + [1-\alpha(Z_i)]\exp(-\mathrm{e}^{\beta_2^T Z_i}\Lambda_2(X_i))},$$

and $V_{2i}(\theta) = 1 - V_{1i}(\theta)$. Then straightforward computation gives

$$\begin{aligned}&\ell_{1,\theta}[h_1](X_1,\triangle_1,Z_1)\\ &= h_1^T\alpha'(Z_1)\left\{\frac{[\triangle_1=1]}{\alpha(Z_1)} - \frac{[\triangle_1=2]}{1-\alpha(Z_1)} + [\triangle_1=3]\left(\frac{V_{11}(\theta)}{\alpha(Z_1)} - \frac{V_{21}(\theta)}{1-\alpha(Z_1)}\right)\right\},\end{aligned} \tag{6}$$

where $\alpha'(Z_1)$ is the derivative of α and is equal to $\alpha(Z_1)(1-\alpha(Z_1))[1, Z_1^T]^T$;

$$\begin{aligned}&\ell_{2,\theta}[h_2](X_1,\triangle_1,Z_1)\\ &= h_2^T Z_1\left\{[\triangle_1=1](1-\Lambda_1(X_1)\mathrm{e}^{\beta_1^T Z_1}) - [\triangle_1=3](\Lambda_1(X_1)\mathrm{e}^{\beta_1^T Z_1}V_{11}(\theta))\right\};\end{aligned} \tag{7}$$

$$\ell_{3,\,\theta}[h_3](X_1,\triangle_1,Z_1)$$
$$=h_3^TZ_1\left\{[\triangle_1=2](1-\Lambda_2(X_1)\,\mathrm{e}^{\beta_2^TZ_1})-[\triangle_1=3](\Lambda_2(X_1)\,\mathrm{e}^{\beta_2^TZ_1}V_{21}(\theta))\right\}; \tag{8}$$

$$\ell_{4,\,\theta}[h_4](X_1,\triangle_1,Z_1)$$
$$=[\triangle_1=1]\left\{h_4(X_1)-\mathrm{e}^{\beta_1^TZ_1}\int_0^{X_1}h_4\,\mathrm{d}\Lambda_1\right\}-[\triangle_1=3]\left\{V_{11}(\theta)\,\mathrm{e}^{\beta_1^TZ_1}\int_0^{X_1}h_4\,\mathrm{d}\Lambda_1\right\}; \tag{9}$$

$$\ell_{5,\theta}[h_5](X_1,\triangle_1,Z_1)$$
$$=[\triangle_1=2]\left\{h_5(X_1)-\mathrm{e}^{\beta_2^TZ_1}\int_0^{X_1}h_5\,\mathrm{d}\Lambda_2\right\}-[\triangle_1=3]\left\{V_{21}(\theta)\,\mathrm{e}^{\beta_2^TZ_1}\int_0^{X_1}h_5\,\mathrm{d}\Lambda_2\right\}. \tag{10}$$

By theorem 1 it is clear that a necessary condition for $\hat\theta_n$ to be the NPMLE is $P_n\ell_{j,\hat\theta_n}[h_j]=0$ for $j=4$ and 5. Here, P_n means taking expectation relative to the empirical distribution for the data $\{(X_i,\triangle_i,Z_i)\mid i=1,\ldots,n\}$; i.e., $P_ng\equiv\frac{1}{n}\sum_{i=1}^n g(X_i,\triangle_i,Z_i)$, for a function g on the range of $(X_i,\triangle_i,Z_i)$. In fact, we will show that the NPMLE converges to the true value almost surely, and hence $\hat\alpha_{cn}$, $\hat\beta_{1n}$, and $\hat\beta_{2n}$ are interior points of $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{B}$, respectively, for large n. This shows that

$$P_n\ell_{j,\hat\theta_n}[h_j]=0, \tag{11}$$

for all large n and $j=1,\ldots,5$.

The statements $(i),(iii),(vii),(viii)$ and (ix) in the following lemma are consequences of (11), and form the basis of our algorithm for computing the NPMLE.

Let $\zeta=(\alpha_c^T,\beta_1^T,\beta_2^T)$ denote a point in $\mathcal{A}\times\mathcal{B}\times\mathcal{B}$. Let $\hat\Lambda_{1n,\,\zeta}$ and $\hat\Lambda_{2n,\,\zeta}$ be the elements in $\mathcal{L}_*$ so that $(\alpha_c^T,\beta_1^T,\beta_2^T,\hat\Lambda_{1n,\,\zeta},\hat\Lambda_{2n,\,\zeta})$ maximizes (5) with ζ being fixed; namely $L_n(\zeta,\hat\Lambda_{1n,\,\zeta},\hat\Lambda_{2n,\,\zeta})\geq L_n(\zeta,\Lambda_1,\Lambda_2)$ for every Λ_1 and Λ_2 in $\mathcal{L}_*$. Both $\Lambda_{1n,\,\zeta}$ and $\Lambda_{2n,\,\zeta}$ play important roles in the asymptotic theory in this paper.

Let

$$W_{1n}(\theta;u)=\frac{1}{n}\sum_{i=1}^n\left\{[\triangle_i=1]+[\triangle_i=3]V_{1i}(\theta)\right\}\mathrm{e}^{\beta_1^TZ_i}I_{(0,\,X_i]}(u),$$
$$W_{2n}(\theta;u)=\frac{1}{n}\sum_{i=1}^n\left\{[\triangle_i=2]+[\triangle_i=3]V_{2i}(\theta)\right\}\mathrm{e}^{\beta_2^TZ_i}I_{(0,\,X_i]}(u),$$

$$G_{1n}(u)=\frac{1}{n}\sum_{i=1}^n[\triangle_i=1]I_{(0,\,u]}(X_i),\quad G_{2n}(u)=\frac{1}{n}\sum_{i=1}^n[\triangle_i=2]I_{(0,\,u]}(X_i),$$
$$W_{10}(\theta;u)=EW_{11}(\theta;u),\quad W_{20}(\theta;u)=EW_{21}(\theta;u),$$
$$G_{10}(u)=EG_{11}(u),\quad G_{20}(u)=EG_{21}(u).$$

Then we have

Lemma 1

(i) $\hat\Lambda_{1n}(t)=\int_0^t\frac{1}{W_{1n}(\hat\theta_n;u)}\,\mathrm{d}G_{1n}(u)$,

(ii) $\Lambda_{10}(t)=\int_0^t\frac{1}{W_{10}(\theta_0;u)}\,\mathrm{d}G_{10}(u)$,

(iii) $\hat\Lambda_{2n}(t)=\int_0^t\frac{1}{W_{2n}(\hat\theta_n;u)}\,\mathrm{d}G_{2n}(u)$,

(iv) $\Lambda_{20}(t)=\int_0^t\frac{1}{W_{20}(\theta_0;u)}\,\mathrm{d}G_{20}(u)$,

$$(v)\ \hat{\Lambda}_{1n,\zeta}(t)=\int_0^t \frac{1}{W_{1n}(\zeta,\hat{\Lambda}_{1n,\zeta},\hat{\Lambda}_{2n,\zeta};u)}\,\mathrm{d}G_{1n}(u),$$

$$(vi)\ \hat{\Lambda}_{2n,\zeta}(t)=\int_0^t \frac{1}{W_{2n}(\zeta,\hat{\Lambda}_{1n,\zeta},\hat{\Lambda}_{2n,\zeta};u)}\,\mathrm{d}G_{2n}(u),$$

$$(vii)\ (\hat{\alpha}_{cn})_j=\log\left(\mathrm{e}^{(\hat{\alpha}_{cn})_j}\frac{\sum_{i=1}^{n}(1-\hat{\alpha}_n(Z_i))([1,Z_i^T]^T)_j([\triangle_i=1]+[\triangle_i=3]V_{1i}(\hat{\theta}_n))}{\sum_{i=1}^{n}\hat{\alpha}_n(Z_i)([1,Z_i^T]^T)_j([\triangle_i=2]+[\triangle_i=3]V_{2i}(\hat{\theta}_n))}\right),\ \textit{for } j=1,\ldots,d+1,$$

$$(viii)\ (\hat{\beta}_{1n})_j=\log\left(\mathrm{e}^{(\hat{\beta}_{1n})_j}\frac{\sum_{i=1}^{n}(Z_i)_j[\triangle_i=1]}{\sum_{i=1}^{n}(Z_i)_j\hat{\Lambda}_{1n}(X_i)\mathrm{e}^{\hat{\beta}_{1n}^T Z_i}\left\{[\Delta_i=1]+[\Delta_i=3]V_{1i}(\hat{\theta}_n)\right\}}\right),\ \textit{for } j=1,\ldots,d,$$

$$(ix)\ (\hat{\beta}_{2n})_j=\log\left(\mathrm{e}^{(\hat{\beta}_{2n})_j}\frac{\sum_{i=1}^{n}(Z_i)_j[\triangle_i=2]}{\sum_{i=1}^{n}(Z_i)_j\hat{\Lambda}_{2n}(X_i)\mathrm{e}^{\hat{\beta}_{2n}^T Z_i}\left\{[\Delta_i=2]+[\Delta_i=3]V_{1i}(\hat{\theta}_n)\right\}}\right),\ \textit{for } j=1,\ldots,d;$$

where $(\cdot)_j$ *is the jth coodinate of the vector.*

Proof. The proofs for (iii), (v) and (vi) are similar to that for (i), and that for (iv) is similar to (ii); thus they are omitted. The proofs for (vii), $(viii)$ and (ix) are straightforward and hence also omitted, but we will provide some heuristics that motivates them at the end of this section. We first give a detailed proof for (i).

Since $P_n\ell_{4,\hat{\theta}_n}[h_4]=0$ for every $h_4\in BV[0,\tau]$, we set $h_4(t)=I_{(0,u]}(t)$ in (9) and get, for every u,

$$\sum_{i=1}^{n}[\triangle_i=1]I_{(0,u]}(X_i)=\sum_{i=1}^{n}\left(\left\{[\triangle_i=1]+[\triangle_i=3]V_{1i}(\hat{\theta}_n)\right\}\mathrm{e}^{\hat{\beta}_{1n}^T Z_i}\int_0^u I_{(0,X_i]}(t)\,\mathrm{d}\hat{\Lambda}_{1n}(t)\right).$$

This leads to $G_{1n}(u)=\int_0^u W_{1n}(\hat{\theta}_n;t)\,\mathrm{d}\hat{\Lambda}_{1n}(t)$, which in turn immediately gives (i).

Since (ii) can be proved similarly by using $E\ell_{4,\theta_0}[h_4](X_1,\triangle_1,Z_1)=0$, for every $h_4\in BV[0,\tau]$, the proof is complete.

Making use of lemma 1, we now present the algorithm for computing the NPMLE. Let $\mathbf{\Lambda}_1(\hat{\theta}_n)(t)$, $\mathbf{\Lambda}_2(\hat{\theta}_n)(t)$, $\mathbf{A}(\hat{\theta}_n)$, $\mathbf{B}_1(\hat{\theta}_n)$, and $\mathbf{B}_2(\hat{\theta}_n)$ denote respectively the right-hand side of (i), (iii), (vii), $(viii)$ and (ix) in lemma 1. The algorithm is given as follows.

(1) Choose starting values $\alpha_c^{(1)}$, $\beta_1^{(1)}$, $\beta_2^{(1)}$, $\Lambda_1^{(1)}$, and $\Lambda_2^{(1)}$.
(2) Set $K=1$.
(3) $\alpha_c^{(K+1)}=\mathbf{A}(\alpha_c^{(K)},\beta_1^{(K)},\beta_2^{(K)},\Lambda_1^{(K)},\Lambda_2^{(K)})$.
(4) $\beta_1^{(K+1)}=\mathbf{B}_1(\alpha_c^{(K+1)},\beta_1^{(K)},\beta_2^{(K)},\Lambda_1^{(K)},\Lambda_2^{(K)})$.
(5) $\beta_2^{(K+1)}=\mathbf{B}_2(\alpha_c^{(K+1)},\beta_1^{(K+1)},\beta_2^{(K)},\Lambda_1^{(K)},\Lambda_2^{(K)})$.
(6) $\Lambda_1^{(K+1)}(t)=\mathbf{\Lambda}_1(\alpha_c^{(K+1)},\beta_1^{(K+1)},\beta_2^{(K+1)},\Lambda_1^{(K)},\Lambda_2^{(K)})(t)$.
(7) $\Lambda_2^{(K+1)}(t)=\mathbf{\Lambda}_2(\alpha_c^{(K+1)},\beta_1^{(K+1)},\beta_2^{(K+1)},\Lambda_1^{(K+1)},\Lambda_2^{(K)})(t)$.
(8) $K=K+1$.
(9) Repeat (3)–(8) for a suitable number M of iterations until there is evidence of convergence.
(10) The estimate of θ is given by $\hat{\theta}_n=(\alpha_c^{(M)},\beta_1^{(M)},\beta_2^{(M)}\Lambda_1^{(M)},\Lambda_2^{(M)})$.

Here are some heuristics for using (vii), $(viii)$ and (ix) in the algorithm. Suppose the score function is the derivative η_1' of a function η_1, and $\eta_1'(\gamma_0)=0$ with $\gamma_0>0$ being the NPMLE of certain parameter. If $\eta_1'=\eta_2-\eta_3$ for two positive functions η_2 and η_3, $\gamma^{(1)}$ is in a suitable neighbourhood of γ_0, and $\gamma^{(K+1)}=\gamma^{(K)}\frac{\eta_2(\gamma^{(K)})}{\eta_3(\gamma^{(K)})}$, then $\gamma^{(K)}$ converges to γ_0 in view of the proposition 3 in Chang *et al.* (2006), given below. With suitably chosen η_2 and η_3, we can get (vii), $(viii)$ and (ix). Consider $(viii)$, for example. We set

$$\eta_2 = \sum_{i=1}^{n} (Z_i)_j [\triangle_i = 1],$$

and

$$\eta_3 = \sum_{i=1}^{n} (Z_i)_j \hat{\Lambda}_1(X_i) \mathrm{e}^{\hat{\beta}_1^T Z_i} \left\{ [\Delta_i = 1] + [\Delta_i = 3] V_{1i}(\hat{\theta}_n) \right\}.$$

Then $\eta_2 - \eta_3$ is the score (7) and the step (4) in the algorithm is an implementation of $\gamma^{(K+1)} = \gamma^{(K)} \frac{\eta_2(\gamma^{(K)})}{\eta_3(\gamma^{(K)})}$.

Proposition (Chang *et al.*, 2006)
Let $\eta_1 : (a, b) \to \Re$ be a function possessing bounded continuous second derivative. Let $\gamma_0 > 0$ be an isolated local maximum of η_1 and $\eta_1''(\gamma_0) < 0$. Let η_2 and η_3 be two positive and continuously differentiable functions satisfying $\eta_1' = \eta_2 - \eta_3$. Then there exist $\varepsilon > 0$ and $n_0 > 0$ such that if $|\gamma_1 - \gamma_0| < \varepsilon$, and $\gamma_{J+1} = \gamma_J \frac{\eta_2(\gamma_J) + n_0}{\eta_3(\gamma_J) + n_0}$ for $J = 1, 2, \ldots,$ then γ_J converges to γ_0.

The idea behind this proposition is simple and goes as follows. If $0 < \gamma_J < \gamma_0$, then we would like to have $\gamma_{J+1} > \gamma_J$. If γ_J is in a suitable neighbourhood of γ_0, then $\eta_2(\gamma_J) - \eta_3(\gamma_J) = \eta_1'(\gamma_J) > 0$, and hence $\gamma_{J+1} > \gamma_J$. Similar comments hold when $\gamma_J > \gamma_0$.

3. Consistency of NPMLE

The purpose of this section is to establish the following theorem 2 and theorem 2′. As much as theorem 2 is an indispensable part of the asymptotic theory, theorem 2′ concerns a useful consistency condition for the profile likelihood theory, discussed in Murphy & van der Vaart (2000). Since the proof for theorem 2 is similar to that for theorem 2′, we prove only the latter.

Theorem 2 (Consistency)
$||\hat{\alpha}_{cn} - \alpha_{c0}||$, $||\hat{\beta}_{1n} - \beta_{10}||$, $||\hat{\beta}_{2n} - \beta_{20}||$, $\sup_{t \in [0, \tau]} |\hat{\Lambda}_{1n}(t) - \Lambda_{10}(t)|$, *and* $\sup_{t \in [0, \tau]} |\hat{\Lambda}_{2n}(t) - \Lambda_{20}(t)|$ *converge to* 0 *almost surely, as n tends to infinity, where* $|| \cdot ||$ *is the Euclidean norm.*

Theorem 2′
For $n = 1, 2, \ldots,$ let $\zeta_n = (\alpha_{cn}, \beta_{1n}, \beta_{2n})$ be a random element in $\mathcal{A} \times \mathcal{B} \times \mathcal{B}$ that converges in probability to $\zeta_0 = (\alpha_{c0}, \beta_{10}, \beta_{20})$. Then both $\sup_{t \in [0, \tau]} |\hat{\Lambda}_{1n, \zeta_n}(t) - \Lambda_{10}(t)|$ *and* $\sup_{t \in [0, \tau]} |\hat{\Lambda}_{2n, \zeta_n}(t) - \Lambda_{20}(t)|$ *converge to* 0 *almost surely as n goes to infinity.*

We need a few lemmas, before presenting the proof.

Lemma 2
Let $\Theta_c = \mathcal{A} \times \mathcal{B} \times \mathcal{B} \times \mathcal{L}_c \times \mathcal{L}_c$, where $\mathcal{L}_c = \{\Lambda \in \mathcal{L} \mid \Lambda(\tau) \le c\}$ for some $c \in (0, \infty)$. Then

$$\sup_{u \in [0, \tau],\, \theta \in \Theta_c} |W_{1n}(\theta; u) - W_{10}(\theta; u)| \quad \textit{and} \quad \sup_{u \in [0, \tau],\, \theta \in \Theta_c} |W_{2n}(\theta; u) - W_{20}(\theta; u)|$$

converge to 0 *almost surely, as n goes to infinity.*

Proof. We will prove the first part of lemma 2, the other goes the same way. Let

$$g_1(\theta, u, X_1, \triangle_1, Z_1) = \{[\triangle_1 = 1] + [\triangle_1 = 3] V_{11}(\theta)\} \mathrm{e}^{\beta_1^T Z_1} I_{(0, X_1]}(u).$$

Then $W_{1n}(\theta; u) - W_{10}(\theta; u) = P_n g_1(\theta, u, \cdot, \cdot, \cdot) - Eg_1(\theta, u, X_1, \triangle_1, Z_1)$. By repeatedly using the theorem 2.10.6 and examples 2.10.4, 2.10.7, 2.10.8, 2.10.9, and 2.10.27 in van der Vaart & Wellner (1996), we know $\mathcal{G}_1 = \{g_1(\theta, u, \cdot, \cdot, \cdot) \mid \theta \in \Theta_c, u \in [0, \tau]\}$ is a Donsker class. This implies that $\sqrt{n}(P_n g_1(\theta, u, \cdot, \cdot, \cdot) - Eg_1(\theta, u, X_1, \delta_1, Z_1))$ converges weakly to a tight Borel measurable Gaussian element in $\ell^\infty(\mathcal{G}_1)$ as n goes to infinity. Thus $P_n g_1(\theta, u, \cdot, \cdot, \cdot) - Eg_1(\theta, u, X_1, \delta_1, Z_1)$ converges weakly to zero as a random element in $\ell^\infty(\mathcal{G}_1)$. This completes the proof.

Lemma 3
Let $\theta \in \Theta_c$ be given. Then

$$\sup_{t\in[0,\tau]} \left| \int_0^t \frac{1}{W_{1n}(\theta; u)} \, \mathrm{d}(G_{1n}(u) - G_{10}(u)) \right| \quad \textit{and} \quad \sup_{t\in[0,\tau]} \left| \int_0^t \frac{1}{W_{2n}(\theta; u)} \, \mathrm{d}(G_{2n}(u) - G_{20}(u)) \right|$$

converge to 0 *almost surely, as n goes to infinity.*

Proof. It follows from the permanence of the Donsker property and the fact that the class of non-negative increasing functions with a common upper bound is Donsker that the class of functions $[\triangle_i = 1] I_{(0,u]}(X_i)$, indexed by u, is Donsker and hence Glivenko-Cantelli (van der Vaart & Wellner, 1996, examples 2.10.4, 2.10.7, and 2.10.8). This shows that $\sup_{t\in[0,\tau]} |G_{1n}(t) - G_{10}(t)|$ goes to zero almost surely. Combining (13), the uniform convergence of G_{1n}, monotonicity of $W_{1n}(\theta; \cdot)$, and the following integration by parts equation

$$\int_0^t \frac{1}{W_{1n}(\theta; u)} \, \mathrm{d}(G_{1n}(u) - G_{10}(u)) = \left. \frac{G_{1n}(u) - G_{10}(u)}{W_{1n}(\theta; u)} \right|_0^t - \int_0^t (G_{1n}(u) - G_{10}(u)) \, \mathrm{d} \frac{1}{W_{1n}(\theta; u)},$$

we get the desired result immediately.

Using lemmas 2 and 3, we show the following.

Lemma 4
Both

$$(i) \sup_{t\in[0,\tau]} \left| \int_0^t \frac{1}{W_{1n}(\theta; u)} \, \mathrm{d}G_{1n}(u) - \int_0^t \frac{1}{W_{10}(\theta; u)} \, \mathrm{d}G_{10}(u) \right|,$$

and

$$(ii) \sup_{t\in[0,\tau]} \left| \int_0^t \frac{1}{W_{2n}(\theta; u)} \, \mathrm{d}G_{2n}(u) - \int_0^t \frac{1}{W_{20}(\theta; u)} \, \mathrm{d}G_{20}(u) \right|$$

converge to 0 *almost surely as n goes to infinity.*

Proof. We prove (*i*) only, because (*ii*) is similar. Consider

$$\begin{aligned} &\sup_{t\in[0,\tau]} \left| \int_0^t \frac{1}{W_{1n}(\theta; u)} \, \mathrm{d}G_{1n}(u) - \int_0^t \frac{1}{W_{10}(\theta; u)} \, \mathrm{d}G_{10}(u) \right| \\ &\leq \sup_{t\in[0,\tau]} \left| \int_0^t \frac{1}{W_{1n}(\theta; u)} \, \mathrm{d}(G_{1n} - G_{10})(u) \right| \\ &\quad + \sup_{t\in[0,\tau]} \left| \int_0^t \left(\frac{1}{W_{1n}(\theta; u)} - \frac{1}{W_{10}(\theta; u)} \right) \mathrm{d}G_{10}(u) \right|. \end{aligned} \tag{12}$$

From the definition of W_{1n}, there exist $c_2 > 0$ and $c_3 > 0$ such that

$$\frac{c_3}{n}\sum_{i=1}^{n}(I_{(0,\,X_i]}(\tau)I_{[\triangle_i=1]}) \le W_{1n}(\theta;u) \le \frac{c_2}{n}\sum_{i=1}^{n}(I_{(0,\,X_i]}(u)I_{[\triangle_i=1\,or\,3]}) \le c_2,$$

for every (θ, u) in $\Theta\times[0,\tau]$; we know from the condition $P(T_1 \ge \tau, W_1 = 1) > 0$ and the Law of Large Numbers that

$$0 < c_1 \le W_{1n}(\theta;u) \le c_2, \tag{13}$$

almost surely for large n. Here $c_1 = c_3P(T_1 \ge \tau, W_1 = 1)$. Using (12), lemmas 2, 3 and (13), we get (i) immediately. This completes the proof.

Lemma 5

$$\limsup_{n\to\infty}\hat{\Lambda}_{1n}(\tau) < \infty, a.s. \quad \textit{and} \quad \limsup_{n\to\infty}\hat{\Lambda}_{2n}(\tau) < \infty, a.s.$$

Lemma 5′

$$\limsup_{n\to\infty}\hat{\Lambda}_{1n,\,\zeta_n}(\tau) < \infty, a.s. \quad \textit{and} \quad \limsup_{n\to\infty}\hat{\Lambda}_{2n,\,\zeta_n}(\tau) < \infty, a.s.$$

We note that lemma 5 and lemma 5′ are, respectively, important steps in proving theorem 2 and theorem 2′, and since their proofs are similar, we prove only lemma 5′.

Proof. We prove only $\limsup_{n\to\infty}\hat{\Lambda}_{1n,\,\zeta_n}(\tau) < \infty$, a.s., because the other is similar. Let

$$\tilde{\Lambda}_{1n,\,\zeta_n}(t) = \int_0^t \frac{1}{W_{1n}(\zeta_n, \Lambda_{10}, \Lambda_{20}; u)}\,\mathrm{d}G_{1n}(u) \tag{14}$$

and

$$\tilde{\Lambda}_{2n,\,\zeta_n}(t) = \int_0^t \frac{1}{W_{2n}(\zeta_n, \Lambda_{10}, \Lambda_{20}; u)}\,\mathrm{d}G_{2n}(u). \tag{15}$$

It follows from lemma 4, (ii) and (iv) of lemma 1 and 2 and the condition ζ_n converging to ζ_0 in probability that both $\sup_{t\in[0,\,\tau]}\left|\tilde{\Lambda}_{1n,\,\zeta_n}(t) - \Lambda_{10}(t)\right|$ and $\sup_{t\in[0,\,\tau]}\left|\tilde{\Lambda}_{2n,\,\zeta_n}(t) - \Lambda_{20}(t)\right|$ converge to 0 almost surely. Let $A_i = [X_i = \tau]$. Since $\sum_i P(A_i) = \infty$ and $\{A_i\}$ are independent, we have $P(A_i\, i.o.) = 1$, by the Borel-Cantelli lemma. Here, the abbreviation '$i.o.$' stands for 'infinitely often'.

Suppose $\limsup_{n\to\infty}\hat{\Lambda}_{1n,\,\zeta_n}(\tau) = \infty$ with positive probability, then there exists an $\omega \in [A_i\, i.o.]$ satisfying $\sup_{t\in[0,\,\tau]}|\tilde{\Lambda}_{1n,\,\zeta_n}(t) - \Lambda_{10}(t)| \to 0$, $\sup_{t\in[0,\,\tau]}|\tilde{\Lambda}_{2n,\,\zeta_n}(t) - \Lambda_{20}(t)| \to 0$, and $\limsup_{n\to\infty}\hat{\Lambda}_{1n,\,\zeta_n}(\tau) = \infty$. Let $\{n_j\}$ be a subsequence such that $\hat{\Lambda}_{1n_j,\,\zeta_{n_j}}(\tau) \to \infty$ for this ω.

Let $\tilde{\theta}_n = (\zeta_n, \tilde{\Lambda}_{1n,\,\zeta_n}, \tilde{\Lambda}_{2n,\,\zeta_n})$ and $\bar{\theta}_n = (\zeta_n, \hat{\Lambda}_{1n,\,\zeta_n}, \hat{\Lambda}_{2n,\,\zeta_n})$. Because $(\hat{\Lambda}_{1n,\,\zeta_n}, \hat{\Lambda}_{2n,\,\zeta_n})$ maximizes $L_{n_j}(\zeta_n, \cdot, \cdot)$, we know

$$\begin{aligned}
0 &\le \frac{1}{n_j}\log L_{n_j}(\bar{\theta}_{n_j}) - \frac{1}{n_j}\log L_{n_j}(\tilde{\theta}_{n_j}) \\
&= \frac{1}{n_j}\sum_{i=1}^{n_j}(I_{A_i} + I_{A_i^c})\log\left\{\frac{\bar{L}_{(i),\,\bar{\theta}_{n_j}}}{\bar{L}_{(i),\,\tilde{\theta}_{n_j}}}\right\} \\
&\le \log\left\{\frac{1}{n_j}\sum_{i=1}^{n_j}\frac{\bar{L}_{(i),\,\bar{\theta}_{n_j}}}{\bar{L}_{(i),\,\tilde{\theta}_{n_j}}}\mathbf{1}_{A_i}\right\} + \log\left\{\frac{1}{n_j}\sum_{i=1}^{n_j}\frac{\bar{L}_{(i),\,\bar{\theta}_{n_j}}}{\bar{L}_{(i),\,\tilde{\theta}_{n_j}}}\mathbf{1}_{A_i^c}\right\}.
\end{aligned} \tag{16}$$

We note that the last inequality of (16) follows from the Jensen's inequality.

Using lemma 1 and (14), we know that

$$\frac{\triangle\hat{\Lambda}_{1n_j,\zeta_{n_j}}(X_i)}{\triangle\tilde{\Lambda}_{1n_j,\zeta_{n_j}}(X_i)} = \frac{W_{1n_j}(\zeta_{n_j},\Lambda_{10},\Lambda_{20};X_i)}{W_{1n_j}(\zeta_{n_j},\hat{\Lambda}_{1n_j,\zeta_{n_j}},\hat{\Lambda}_{2n_j,\zeta_{n_j}};X_i)}. \tag{17}$$

It follows from (13) and (17) that

$$0 < \frac{c_1}{c_2} \le \varliminf_{j\to\infty} \frac{\triangle\hat{\Lambda}_{1n_j,\zeta_{n_j}}(X_i)}{\triangle\tilde{\Lambda}_{1n_j,\zeta_{n_j}}(X_i)} \le \varlimsup_{j\to\infty} \frac{\triangle\hat{\Lambda}_{1n_j,\zeta_{n_j}}(X_i)}{\triangle\tilde{\Lambda}_{n_j,\zeta_{n_j}}(X_i)} \le \frac{c_2}{c_1}. \tag{18}$$

Similarly, there exist $b_1 > 0$ and $b_2 > 0$ such that

$$b_1 \le \varliminf_{j\to\infty} \frac{\triangle\hat{\Lambda}_{2n_j,\zeta_{n_j}}(X_i)}{\triangle\tilde{\Lambda}_{2n_j,\zeta_{n_j}}(X_i)} \le \varlimsup_{j\to\infty} \frac{\triangle\hat{\Lambda}_{2n_j,\zeta_{n_j}}(X_i)}{\triangle\tilde{\Lambda}_{n_j,\zeta_{n_j}}(X_i)} \le b_2. \tag{19}$$

Using (18), (19), and the facts that $(\alpha_{cn_j},\beta_{1n_j},\beta_{2n_j}) \to (\alpha_{c0},\beta_{10},\beta_{20})$, that $\tilde{\Lambda}_{1n_j,\zeta_{n_j}} \to \Lambda_{10}$, $\tilde{\Lambda}_{2n_j,\zeta_{n_j}} \to \Lambda_{20}$ on $[0,\tau]$, and that $n_j^{-1}\sum_{i=1}^{n_j}\exp(-\hat{\Lambda}_{1n_j,\zeta_{n_j}}(X_i))I_{A_i} \to 0$ as $j\to\infty$, we can show that the right hand of (16) at that ω goes to $-\infty$ as $j\to\infty$. This leads to a contradiction, hence $\limsup_{n\to\infty}\hat{\Lambda}_{1n,\zeta_n}(\tau) < \infty$ a.s. This completes the proof.

Proof of theorem 2′. It follows from lemma 5′, arguments in the proof of lemma 5′ and the Law of Large Numbers that there exists $\omega\in\Omega$ for which $\limsup_{n\to\infty}\hat{\Lambda}_{1n,\zeta_n}(\tau)<\infty$, $\limsup_{n\to\infty}\hat{\Lambda}_{2n,\zeta_n}(\tau)<\infty$, $\sup_{t\in[0,\tau]}|G_{1n}(t)-G_{10}(t)|\to 0$, $\sup_{t\in[0,\tau]}|G_{2n}(t)-G_{20}(t)|\to 0$, $\sup_{t\in[0,\tau]}|\tilde{\Lambda}_{1n,\zeta_n}(t)-\Lambda_{10}(t)|\to 0$, and $\sup_{t\in[0,\tau]}|\tilde{\Lambda}_{2n,\zeta_n}(t)-\Lambda_{20}(t)|\to 0$. Here $\tilde{\Lambda}_{1n,\zeta_n}$ is given by (14), and $\tilde{\Lambda}_{2n,\zeta_n}$ is given by (15). Using lemma 1, (13), and the Law of Large Numbers, we know that

$$|\hat{\Lambda}_{1n,\zeta_n}(s)-\hat{\Lambda}_{1n,\zeta_n}(t)| \le O(1)|G_{1n}(s)-G_{1n}(t)| \le O(1)|G_{10}(s)-G_{10}(t)| + o(1)$$

for $s,t\in[0,\tau]$. This together with the compactness of $\mathcal{A}\times\mathcal{B}\times\mathcal{B}$ and the arguments in proving the Arzela–Ascoli theorem (see, for example, theorem 7.25 in Rudin, 1976) implies that there exists a subsequence $\{n_j\}$ for which $(\alpha_{cn_j}^T,\beta_{1n_j}^T,\beta_{2n_j}^T,\hat{\Lambda}_{1n_j,\zeta_{n_j}},\hat{\Lambda}_{2n_j,\zeta_{n_j}})$ converges uniformly to some $\theta^*=(\alpha_{c0}^T,\beta_{10}^T,\beta_{20}^T,\Lambda_1^*,\Lambda_2^*)$. We will show that $\theta^*=\theta_0$.

We now explain that

$$\begin{aligned}
\hat{\Lambda}_{1n_j,\zeta_{n_j}}(t) &= \int_0^t \frac{1}{W_{1n_j}(\zeta_{n_j},\hat{\Lambda}_{1n_j,\zeta_{n_j}},\hat{\Lambda}_{2n_j,\zeta_{n_j}};u)}\,\mathrm{d}G_{1n_j}(u)\\
&= \int_0^t \frac{1}{W_{1n_j}(\zeta_{n_j},\hat{\Lambda}_{1n_j,\zeta_{n_j}},\hat{\Lambda}_{2n_j,\zeta_{n_j}};u)}\,\mathrm{d}G_{10}(u)+o(1)\\
&= \int_0^t \frac{1}{W_{10}(\zeta_{n_j},\hat{\Lambda}_{1n_j,\zeta_{n_j}},\hat{\Lambda}_{2n_j,\zeta_{n_j}};u)}\,\mathrm{d}G_{10}(u)+o(1)\\
&= \int_0^t \frac{1}{W_{10}(\theta^*;u)}\,\mathrm{d}G_{10}(u)+o(1).
\end{aligned}$$

The first equality follows from lemma 1, the second equality can be proved by using lemma 2 and the arguments in proving lemma 3, the third equality follows from lemma 2, and the last equality follows from the uniform convergence of $(\zeta_{n_j},\hat{\Lambda}_{1n_j,\zeta_{n_j}},\hat{\Lambda}_{2n_j,\zeta_{n_j}})$ and the lebesgue dominated convergence theorem. Thus, $\Lambda_1^*(t)=\int_0^t \frac{1}{W_{10}(\theta^*;u)}\,\mathrm{d}G_{10}(u)$. This, together with lemmas 1 and 2, implies, uniformly in t, that

$$\frac{\mathrm{d}\hat{\Lambda}_{1n_j,\zeta_{n_j}}}{\mathrm{d}\tilde{\Lambda}_{1n_j,\zeta_{n_j}}}(t)=\frac{W_{1n_j}(\zeta_{n_j},\Lambda_{10},\Lambda_{20};t)}{W_{1n_j}(\zeta_{n_j},\hat{\Lambda}_{1n_j,\zeta_{n_j}},\hat{\Lambda}_{2n_j,\zeta_{n_j}};t)}\rightarrow\frac{W_{10}(\theta_0,t)}{W_{10}(\theta^*;t)}=\frac{\mathrm{d}\Lambda_1^*}{\mathrm{d}\Lambda_{10}}(t),$$

which is bounded and bounded away from 0 by (13).

Considering $\left[\frac{\mathrm{d}\hat{\Lambda}_{1n_j,\zeta_{n_j}}}{\mathrm{d}\tilde{\Lambda}_{1n_j,\zeta_{n_j}}}(x)\right]^{[\delta=1]}$ as a function in (x,δ) with ω fixed, we can use the argument in the proof of lemma 3 and properties concerning random functions and Glivenko–Cantelli class (see, for example, van der Vaart, 1998, p. 279) to obtain

$$\begin{aligned}&\frac{1}{n_j}\sum_{i=1}^{n_j}\log\left\{\left[\frac{\mathrm{d}\hat{\Lambda}_{1n_j,\zeta_{n_j}}}{\mathrm{d}\tilde{\Lambda}_{1n_j,\zeta_{n_j}}}(X_i)\right]^{[\triangle_i=1]}\right\}-E\log\left\{\left[\frac{\mathrm{d}\hat{\Lambda}_{1n_j,\zeta_{n_j}}}{\mathrm{d}\tilde{\Lambda}_{1n_j,\zeta_{n_j}}}(X_1)\right]^{[\triangle_1=1]}\right\}\\&=\int_0^{\tau}\log\frac{\mathrm{d}\hat{\Lambda}_{1n_j,\zeta_{n_j}}}{\mathrm{d}\tilde{\Lambda}_{1n_j,\zeta_{n_j}}}(t)\,\mathrm{d}\Big(G_{1n_j}(t)-G_{10}(t)\Big)\\&=o(1),\ \text{a.s.}\end{aligned}\tag{20}$$

Similarly,

$$\frac{1}{n_j}\sum_{i=1}^{n_j}\log\left\{\left[\frac{\mathrm{d}\hat{\Lambda}_{2n_j,\zeta_{n_j}}}{\mathrm{d}\tilde{\Lambda}_{2n_j,\zeta_{n_j}}}(X_i)\right]^{[\triangle_i=2]}\right\}-E\log\left\{\left[\frac{\mathrm{d}\hat{\Lambda}_{2n_j,\zeta_{n_j}}}{\mathrm{d}\tilde{\Lambda}_{2n_j,\zeta_{n_j}}}(X_1)\right]^{[\triangle_1=2]}\right\}=o(1),\ \text{a.s.}\tag{21}$$

In fact, the expectations in (20) and (21) are taken only over (X_1,Δ_1) with parameter estimators $\hat{\Lambda}_{1n_j,\zeta_{n_j}}$ and $\hat{\Lambda}_{2n_j,\zeta_{n_j}}$ substituted after taking the expectation.

Because the set of functions

$$\begin{aligned}&g(\theta;X_1,\Delta_1,Z_1)\\&\equiv\log\left\{\{\alpha(Z_1)\,\mathrm{e}^{\beta_1^TZ_1-\Lambda_1(X_1)\mathrm{e}^{\beta_1^TZ_1}}\}^{[\triangle_1=1]}\{[1-\alpha(Z_1)]\,\mathrm{e}^{\beta_2^TZ_1-\Lambda_2(X_1)\mathrm{e}^{\beta_2^TZ_1}}\}^{[\triangle_1=2]}\right.\\&\quad\left.\times\{\alpha(Z_1)\,\mathrm{e}^{-\Lambda_1(X_1)\mathrm{e}^{\beta_1^TZ_1}}+[1-\alpha(Z_1)]\,\mathrm{e}^{-\Lambda_2(X_1)\mathrm{e}^{\beta_2^TZ_1}}\}^{[\triangle_1=3]}\right\},\end{aligned}$$

indexed by $\theta\in\Theta_c$, is Glivenko–Cantelli, we can conclude that

$$\begin{aligned}&\Big|P_{n_j}\Big(g(\zeta_{n_j},\hat{\Lambda}_{1n_j,\zeta_{n_j}},\hat{\Lambda}_{2n_j,\zeta_{n_j}};\cdot,\cdot,\cdot)-g(\zeta_{n_j},\tilde{\Lambda}_{1n_j,\zeta_{n_j}},\tilde{\Lambda}_{2n_j,\zeta_{n_j}};\cdot,\cdot,\cdot)\Big)\\&-E\Big(g(\zeta_{n_j},\hat{\Lambda}_{1n_j,\zeta_{n_j}},\hat{\Lambda}_{2n_j,\zeta_{n_j}};X_1,\Delta_1,Z_1)-g(\zeta_{n_j},\tilde{\Lambda}_{1n_j,\zeta_{n_j}},\tilde{\Lambda}_{2n_j,\zeta_{n_j}};X_1,\Delta_1,Z_1)\Big)\Big|=o(1),\text{a.s.}\end{aligned}\tag{22}$$

Here we use the same arguments concerning random functions in deriving (20). In particular, the expectation in (22) is taken only over (X_1,Δ_1,Z_1) with parameter estimators $\hat{\theta}_{n_j}$ substituted after taking the expectation. The expectation in (23) below is to be understood in the same manner.

Using (20), (21) and (22), we get

$$\begin{aligned}0&\le\frac{1}{n_j}\log L_{n_j}(\zeta_{n_j},\hat{\Lambda}_{1n_j,\zeta_{n_j}},\hat{\Lambda}_{2n_j,\zeta_{n_j}})-\frac{1}{n_j}\log L_{n_j}(\zeta_{n_j},\tilde{\Lambda}_{1n_j,\zeta_{n_j}},\tilde{\Lambda}_{2n_j,\zeta_{n_j}})\\&=\frac{1}{n_j}\sum_{i=1}^{n_j}\log\left(\frac{\mathrm{d}\hat{\Lambda}_{1n_j,\zeta_{n_j}}}{\mathrm{d}\tilde{\Lambda}_{1n_j,\zeta_{n_j}}}(X_i)\right)^{[\triangle_i=1]}+\frac{1}{n_j}\sum_{i=1}^{n_j}\log\left(\frac{\mathrm{d}\hat{\Lambda}_{2n_j,\zeta_{n_j}}}{\mathrm{d}\tilde{\Lambda}_{2n_j,\zeta_{n_j}}}(X_i)\right)^{[\triangle_i=2]}\\&\quad+P_{n_j}\{g(\zeta_{n_j},\hat{\Lambda}_{1n_j,\zeta_{n_j}},\hat{\Lambda}_{2n_j,\zeta_{n_j}};\cdot,\cdot,\cdot)-g(\zeta_{n_j},\tilde{\Lambda}_{1n_j,\zeta_{n_j}},\tilde{\Lambda}_{2n_j,\zeta_{n_j}};\cdot,\cdot,\cdot)\}\\&=E\log\frac{\tilde{L}_{(1),(\zeta_{n_j},\hat{\Lambda}_{1n_j,\zeta_{n_j}},\hat{\Lambda}_{2n_j,\zeta_{n_j}})}}{\tilde{L}_{(1),(\zeta_{n_j},\tilde{\Lambda}_{1n_j,\zeta_{n_j}},\tilde{\Lambda}_{2n_j,\zeta_{n_j}})}}+o(1),\quad\text{a.s.}\end{aligned}\tag{23}$$

Considering the ω chosen at the beginning of the proof for the which (23) is also satisfied, using the Jensen's inequality and the Kullback–Leibler divergence theorem (see, for example, van der Vaart, 1998, p. 62), we get $\tilde{L}_{(1),\,\theta^*} = \tilde{L}_{(1),\,\theta_0}$. Since Λ_1^* is absolutely continuous relative to G_{10} and thus also to Λ_{10} by (*ii*) of lemma 1. Similarly, Λ_2^* is absolutely continuous relative to Λ_{20}. We can now use the identifiability assumption 1 to conclude that $\theta^* = \theta_0$. This completes the proof.

4. Asymptotic normality and rate of convergence

We will prove the asymptotic normality of the NPMLE and the rate of convergence of the non-parameteric part by verifying the conditions in theorem 3.3.1 and lemma 3.3.5 of van der Vaart & Wellner (1996) and in theorem 3.1 of Murphy & van der Vaart (1999). For this, a few lemmas are needed. Let $\mathcal{H} = \Re^{d+1} \times \Re^d \times \Re^d \times BV[0,\tau] \times BV[0,\tau]$. For $\mathbf{h} = (h_1, h_2, h_3, h_4, h_5) \in \mathcal{H}$, we introduce the norm $||\mathbf{h}||_{\mathcal{H}} = ||h_1|| + ||h_2|| + ||h_3|| + ||h_4||_V + ||h_5||_V$. Here $||h||_V$ denotes the sum of the absolute value of $h(0)$ and the total variation of h on $[0,\tau]$ for every $h \in BV[0,\tau]$. Let H_p be the subset of $\mathcal{H}$ with $||\mathbf{h}||_{\mathcal{H}} \le p$ if $p < \infty$. If $p = \infty$, then the previous inequality is strict. Define $\theta(\mathbf{h}) = h_1^T \alpha_c + h_2^T \beta_1 + h_3^T \beta_2 + \int_0^\tau h_4 \, d\Lambda_1 + \int_0^\tau h_5 \, d\Lambda_2$ and consider the parameter space Θ a subset of $\ell^\infty(H_p)$, the space of all bounded real-valued functions on H_p under the supremum norm $||\theta||_{\ell^\infty(H_p)} = \sup_{\mathbf{h} \in H_p} |\theta(\mathbf{h})|$. We note that $(p/\sqrt{d+1})(||\alpha_c - \alpha_{c0}|| \vee ||\beta_1 - \beta_{10}|| \vee ||\beta_2 - \beta_{20}|| \vee ||\Lambda_1 - \Lambda_{10}||_* \vee ||\Lambda_2 - \Lambda_{20}||_*) \le ||\theta - \theta_0||_{\ell^\infty(H_p)} \le 5p(||\alpha_c - \alpha_{c0}|| \vee ||\beta_1 - \beta_{10}|| \vee ||\beta_2 - \beta_{20}|| \vee ||\Lambda_1 - \Lambda_{10}||_* \vee ||\Lambda_2 - \Lambda_{20}||_*)$, where $||\Lambda||_* = \sup_{||h||_V \le 1} |\int_0^\tau h \, d\Lambda|$ is the natural norm for a bounded linear operator on the normed space $BV[0,\tau]$.

Let $\phi_{\theta,\,\mathbf{h}} = \sum_{j=1}^5 \ell_{j,\,\theta}[h_j]$, Define $\Psi_n, \Psi : \Theta \to \ell^\infty(H_p)$ by

$$\Psi_n(\theta)(\mathbf{h}) = P_n \phi_{\theta,\,\mathbf{h}} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^5 \ell_{j,\,\theta}[h_j](X_i, \Delta_i, Z_i),$$

$$\Psi(\theta)(\mathbf{h}) = E\Psi_1(\theta)(\mathbf{h}).$$

Lemma 6
$\sqrt{n}\left(\Psi_n(\theta_0) - \Psi(\theta_0)\right)$ *converges weakly to a Gaussian process* $\mathcal{W}$ *in* $\ell^\infty(H_p)$ *for every* $0 < p < \infty$.

Proof. According to empirical process theory, it is sufficient to show that $\{\phi_{\theta_0,\mathbf{h}} \,|\, \mathbf{h} \in H_p\}$ is a Donsker class. Since the fact that $\{\phi_{\theta_0,\mathbf{h}} \,|\, \mathbf{h} \in H_p\}$ is Donsker can be shown in exactly the same way as that for lemma 7, we omit it. This completes the proof.

Lemma 7
$\{\phi_{\theta,\mathbf{h}} - \phi_{\theta_0,\mathbf{h}} \,|\, ||\theta - \theta_0||_{\ell^\infty(H_p)} < \delta, \mathbf{h} \in H_p\}$ *is Donsker.*

Proof. Because the class of functions with a common upper bound of their total variations is Donsker (see, for example, van der Vaart, 1998, example 19.11), we know $\{h \,|\, h \in BV[0,\tau], ||h||_V < p\}$, $\{\int_0^\cdot h \, d\Lambda_{10} \,|\, h \in BV[0,\tau], ||h||_V < p\}$, and $\{\int_0^\cdot h \, d\Lambda_{20} \,|\, h \in BV[0,\tau], ||h||_V < p\}$ are Donsker classes. In fact, they are uniformly bounded Donsker classes. Using this, (6)–(10), and the fact that product and sum of uniformly bounded Donsker classes are again Donsker classes (see, for example, theorem 2.10.6 in van der Vaart & Wellner, 1996), we get the lemma. This completes the proof.

Lemma 8
$\lim_{\theta\to\theta_0} \sup_{\mathbf{h}\in H_p} E(\phi_{\theta,\mathbf{h}} - \phi_{\theta_0,\mathbf{h}})^2 = 0.$

Proof. Using (6)–(10), we have

$$
\begin{aligned}
&\phi_{\theta,\mathbf{h}}(X_1,\Delta_1,Z_1) - \phi_{\theta_0,\mathbf{h}}(X_1,\Delta_1,Z_1)\\
&\quad = \eta\left(\alpha_c - \alpha_{c0}, \beta_1 - \beta_{10}, \beta_2 - \beta_{20}, h_1, h_2, h_3, (\Lambda_1 - \Lambda_{10})(X_1), \int_0^{X_1} h_4\,\mathrm{d}(\Lambda_1 - \Lambda_{10}),\right.\\
&\qquad \left.(\Lambda_2 - \Lambda_{20})(X_1), \int_0^{X_1} h_5\,\mathrm{d}(\Lambda_2 - \Lambda_{20})\right) + O\left(||\alpha_c - \alpha_{c0}||^2 + \sum_{i=1}^2 ||\beta_i - \beta_{i0}||^2\right.\\
&\qquad \left. + \sum_{i=1}^2 |(\Lambda_i - \Lambda_{i0})(X_1)|^2 + \left|\int_0^{X_1} h_4\,\mathrm{d}(\Lambda_1 - \Lambda_{10})\right|^2 + \left|\int_0^{X_1} h_5\,\mathrm{d}(\Lambda_2 - \Lambda_{20})\right|^2\right)\\
&\quad = \eta\left(\alpha_c - \alpha_{c0}, \beta_1 - \beta_{10}, \beta_2 - \beta_{20}, h_1, h_2, h_3, (\Lambda_1 - \Lambda_{10})(X_1), \int_0^{X_1} h_4\,\mathrm{d}(\Lambda_1 - \Lambda_{10}),\right.\\
&\qquad \left.(\Lambda_2 - \Lambda_{20})(X_1), \int_0^{X_1} h_5\,\mathrm{d}(\Lambda_2 - \Lambda_{20})\right) + \circ\left(||\theta - \theta_0||_{\ell^\infty(H_p)}\right),
\end{aligned}
\tag{24}
$$

for some multi-linear function η, as θ gets close to θ_0. We note that the first equality is the first-order Taylor expansion of $\phi_{\theta,\mathbf{h}}(X_1,\Delta_1,Z_1)$, as a function of $(\alpha_c, \beta_1, \beta_2, \Lambda_1(X_1), \Lambda_2(X_1), \int_0^{X_1} h_4\,\mathrm{d}\Lambda_1, \int_0^{X_1} h_5\,\mathrm{d}\Lambda_2)$; the second follows from the inequality in the first paragraph of this section. Thus lemma 8 follows from (24) immediately.

Let $\mathrm{lin}\Theta$ denote the set of all finite linear combinations of $\theta - \theta_0$, for $\theta \in \Theta$.

Lemma 9
Let $p < \infty$. There is a continuous linear map $\dot{\Psi}_{\theta_0} : \mathrm{lin}\Theta \to \ell^\infty(H_p)$ satisfying

$$||\Psi(\theta) - \Psi(\theta_0) - \dot{\Psi}_{\theta_0}(\theta - \theta_0)||_{\ell^\infty(H_p)} = \circ(||\theta - \theta_0||_{\ell^\infty(H_p)}).$$

In addition, $\dot{\Psi}_{\theta_0}$ has a continuous inverse on its range.

Proof. We first prove the existence of $\dot{\Psi}_{\theta_0}$. It follows from (24) that

$$
\begin{aligned}
\Psi(\theta)(\mathbf{h}) - \Psi(\theta_0)(\mathbf{h}) &= E\left[\phi_{\theta,\mathbf{h}}(X_1,\Delta_1,Z_1) - \phi_{\theta_0,\mathbf{h}}(X_1,\Delta_1,Z_1)\right]\\
&= -\left[\sigma_1(\mathbf{h})^T(\alpha_c - \alpha_{c0}) + \sigma_2(\mathbf{h})^T(\beta_1 - \beta_{10}) + \sigma_3(\mathbf{h})^T(\beta_2 - \beta_{20})\right.\\
&\qquad \left. + \int_0^\tau \sigma_4(\mathbf{h})\,\mathrm{d}(\Lambda_1 - \Lambda_{10}) + \int_0^\tau \sigma_5(\mathbf{h})\,\mathrm{d}(\Lambda_2 - \Lambda_{20})\right] + R(\theta)(\mathbf{h}),
\end{aligned}
$$

for some continuous linear operator $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5)$ from H_∞ to H_∞, and remainder term $R(\theta)$ satisfying

$$\lim_{\theta\to\theta_0} \frac{\sup_{\mathbf{h}\in H_p} |R(\theta)(\mathbf{h})|}{||\theta - \theta_0||_{\ell^\infty(H_p)}} = 0.$$

In fact,

$$
\begin{aligned}
\dot{\Psi}_{\theta_0}(\theta - \theta_0)(\mathbf{h}) = &-\left[\sigma_1(\mathbf{h})^T(\alpha_c - \alpha_{c0}) + \sigma_2(\mathbf{h})^T(\beta_1 - \beta_{10}) + \sigma_3(\mathbf{h})^T(\beta_2 - \beta_{20})\right.\\
&\left. + \int_0^\tau \sigma_4(\mathbf{h})\,\mathrm{d}(\Lambda_1 - \Lambda_{10}) + \int_0^\tau \sigma_5(\mathbf{h})\,\mathrm{d}(\Lambda_2 - \Lambda_{20})\right].
\end{aligned}
\tag{25}
$$

With the existence of $\dot{\Psi}_{\theta_0}$, we can prove its invertibility by the arguments in the lemma 4.4 of Chang *et al.* (2005), using the following lemmas 10 and 11 on σ. This completes the proof.

We note that σ is called the information operator. Considering (25) and the negative second directional derivative of the log-likelihood of the parametric submodel

$$(\varepsilon_1, \varepsilon_2) \mapsto (\alpha_{c0} + \varepsilon_1 h_1 + \varepsilon_2 h_1^*, \beta_{10} + \varepsilon_1 h_2 + \varepsilon_2 h_2^*, \beta_{20} + \varepsilon_1 h_3 + \varepsilon_2 h_3^*,$$
$$\Lambda_{10} + \varepsilon_1 \int_0^{\cdot} h_4 \,\mathrm{d}\Lambda_{10} + \varepsilon_2 \int_0^{\cdot} h_4^* \,\mathrm{d}\Lambda_{10}, \Lambda_{20} + \varepsilon_1 \int_0^{\cdot} h_5 \,\mathrm{d}\Lambda_{20} + \varepsilon_2 \int_0^{\cdot} h_5^* \,\mathrm{d}\Lambda_{20})$$

for $\varepsilon_1, \varepsilon_2$ near 0, we get the following equation connecting the information and the score:

$$h_1^{*T}\sigma_1(\mathbf{h}) + h_2^{*T}\sigma_2(\mathbf{h}) + h_3^{*T}\sigma_3(\mathbf{h}) + \int_0^{\tau} \sigma_4(\mathbf{h})(u) h_4^*(u) \,\mathrm{d}\Lambda_{10}(u) + \int_0^{\tau} \sigma_5(\mathbf{h})(u) h_5^*(u) \,\mathrm{d}\Lambda_{20}(u)$$
$$= E\left[\sum_{j=1}^{5} \ell_{j,\theta_0}[h_j](X_1, \Delta_1, Z_1)\right]\left[\sum_{j=1}^{5} \ell_{j,\theta_0}[h_j^*](X_1, \Delta_1, Z_1)\right], \tag{26}$$

for $(h_1, h_2, h_3, h_4, h_5)$ and $(h_1^*, h_2^*, h_3^*, h_4^*, h_5^*)$ in H_∞.

Lemma 10
σ is one to one.

Proof. Assume $\sigma(\mathbf{h}) = 0$. Using (26), we know

$$\sum_{j=1}^{5} \ell_{j,\theta_0}[h_j](X_1, \Delta_1, Z_1) = 0, \quad \text{a.s.} \tag{27}$$

Considering X_1 near t^* from the right in (27) and assumption 2′ with $(a_1, a_2, a_3, a_4, a_5, a_6, a_7) = (h_1, h_2, h_3, \int_0^{t^*} h_4 \,\mathrm{d}\Lambda_{10}, \int_0^{t^*} h_5 \,\mathrm{d}\Lambda_{20}, h_4(t^*), h_5(t^*))$, we know $h_1 = 0, h_2 = h_3 = 0$, and $\int_0^{t^*} h_4 \,\mathrm{d}\Lambda_{10} = \int_0^{t^*} h_5 \,\mathrm{d}\Lambda_{20} = h_4(t^*) = h_5(t^*) = 0$.

Putting $h_1 = 0, h_2 = h_3 = 0$ in (27) and considering $\Delta_1 = 1$, we get

$$h_4(X_1) = \mathrm{e}^{\beta_{10}^T Z_1} \int_0^{X_1} h_4 \,\mathrm{d}\Lambda_{10}$$

for almost every (X_1, Z_1). Let z be any point in the support of the distribution of Z_1. Define $g(t) = \mathrm{e}^{\beta_{10}^T z} \int_0^t h_4 \,\mathrm{d}\Lambda_{10}$, then $g = h_4$ a.s. $[\Lambda_{10}]$. It's easy to show that $g(t) = b\,\mathrm{e}^{\beta_{10}^T z} \exp(\Lambda_{10}(t)\,\mathrm{e}^{\beta_{10}^T z})$ for some constant b. By $g(0) = 0$, we have $g = 0$ identically and hence $h_4 = 0$ a.s. $[\Lambda_{10}]$. Similarly, we have $h_5 = 0$ a.s. $[\Lambda_{20}]$.

Putting $h_1 = 0, h_2 = h_3 = 0, h_4 = 0$ a.s. $[\Lambda_{10}]$, and $h_5 = 0$ a.s. $[\Lambda_{20}]$ in $\sigma_4(\mathbf{h}) = 0$, we obtain from (28) below that $h_4(u) W_{10}(\theta_0; u) = 0$ for $u \in [0, \tau]$. Since $W_{10}(\theta_0; \cdot)$ is uniformly bounded away from zero, $h_4 = 0$ identically. A similar argument gives $h_5 = 0$ identically. This completes the proof.

Lemma 11
σ is continuously invertible.

Proof. Define $A: H_\infty \to H_\infty$ by

$$A(h_1, h_2, h_3, h_4, h_5) = (h_1, h_2, h_3, h_4(\cdot) W_{10}(\theta_0; \cdot), h_5(\cdot) W_{20}(\theta_0; \cdot)).$$

Since $W_{10}(\theta_0,\cdot)$ and $W_{20}(\theta_0,\cdot)$ are uniformly bounded and bounded away from 0 on $[0,\tau]$, A is continuous, linear, and invertible.

Let $\mathcal{K}=\sigma-A$. Hence $\sigma=A(I+A^{-1}\mathcal{K})$. It is sufficient to show that $I+A^{-1}\mathcal{K}$ is continuously invertible. According to lemma 10, $I+A^{-1}\mathcal{K}$ is one to one. This, together with the theorem 4.25 in Rudin (1973), implies that $I+A^{-1}\mathcal{K}$ is invertible, if $A^{-1}\mathcal{K}$ is compact. For this, we need to show that $\mathcal{K}$ is compact.

Let $A=(A_1,A_2,A_3,A_4,A_5)$ and $\mathcal{K}=(\mathcal{K}_1,\mathcal{K}_2,\mathcal{K}_3,\mathcal{K}_4,\mathcal{K}_5)$. We only show $\mathcal{K}_4$ is compact, since the compactness of $\mathcal{K}_1$, $\mathcal{K}_2$, $\mathcal{K}_3$, and $\mathcal{K}_5$ can be shown similarly. Since a bounded linear operator with finite-dimensional range is compact, we need only show that $\mathcal{K}_4$ is compact on $\{(0,0,0,h_4,h_5)\,|\,h_4,h_5\in BV[0,\tau]\}$.

Observing from (25) that

$$
\begin{aligned}
&-\int_0^\tau \sigma_4(0,0,0,h_4,h_5)\,\mathrm{d}(\Lambda_1-\Lambda_{10})\\
&=\dot{\Psi}_{\theta_0}(0,0,0,\Lambda_1-\Lambda_{10},0)(0,0,0,h_4,h_5)\\
&=\left.\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0}\Psi(\alpha_{c0},\beta_{10},\beta_{20},\Lambda_{10}+\varepsilon(\Lambda_1-\Lambda_{10}),\Lambda_{20})(0,0,0,h_4,h_5),
\end{aligned}
$$

we can show that

$$
\begin{aligned}
&\sigma_4(0,0,0,h_4,h_5)(u)\\
&=h_4(u)W_{10}(\theta_0;u)-E[\Delta_1=3]\left\{V_{11}(\theta_0)\,\mathrm{e}^{2\beta_{10}^TZ_1}\int_0^{X_1}h_4\,\mathrm{d}\Lambda_{10}I_{[u\le X_1]}\right.\\
&\quad\left.-\left[V_{11}(\theta_0)\,\mathrm{e}^{\beta_{10}^TZ_1}I_{[u\le X_1]}\right]\left[V_{11}(\theta_0)\,\mathrm{e}^{\beta_{10}^TZ_1}\int_0^{X_1}h_4\,\mathrm{d}\Lambda_{10}+V_{21}(\theta_0)\,\mathrm{e}^{\beta_{20}^TZ_1}\int_0^{X_1}h_5\,\mathrm{d}\Lambda_{20}\right]\right\},
\end{aligned}
\tag{28}
$$

which implies immediately

$$
\begin{aligned}
||\mathcal{K}_4(0,0,0,h_4,h_5)||_V&=||(\sigma_4-A_4)(0,0,0,h_4,h_5)||_V\\
&\le c\left(\int_0^\tau|h_4|\,\mathrm{d}\Lambda_{10}(t)+\int_0^\tau|h_5|\,\mathrm{d}\Lambda_{20}(t)\right),
\end{aligned}
$$

for some constant c. This shows $\mathcal{K}_4$ is compact on $\{(0,0,0,h_4,h_5)\,|\,h_4,h_5\in BV[0,\tau]\}$ by Helly's lemma. This completes the proof.

Theorem 3

$\sqrt{n}(\hat{\theta}_n-\theta_0)$ converges weakly to a tight Gaussian process $\mathcal{G}\equiv-\dot{\Psi}_{\theta_0}^{-1}\mathcal{W}$ on $\ell^\infty(H_p)$ with mean zero and covariance process

$$
\mathrm{Cov}(\mathcal{G}(\boldsymbol{h}),\mathcal{G}(\tilde{\boldsymbol{h}}))=h_1^T\tilde{\sigma}_1(\tilde{\boldsymbol{h}})+h_2^T\tilde{\sigma}_2(\tilde{\boldsymbol{h}})+h_3^T\tilde{\sigma}_3(\tilde{\boldsymbol{h}})+\int_0^\tau h_4\tilde{\sigma}_4(\tilde{\boldsymbol{h}})\,\mathrm{d}\Lambda_{10}+\int_0^\tau h_5\tilde{\sigma}_5(\tilde{\boldsymbol{h}})\,\mathrm{d}\Lambda_{20},\tag{29}
$$

where $(\tilde{\sigma}_1,\tilde{\sigma}_2,\tilde{\sigma}_3,\tilde{\sigma}_4,\tilde{\sigma}_5)=\tilde{\sigma}:H_\infty\mapsto H_\infty$ is the inverse of σ.

Proof of theorem 3. Since lemmas 6, 7, 8 and 9 combined indicate that the conditions in the theorem 3.3.1 and lemma 3.3.5 in van der Vaart & Wellner (1996) are satisfied, we obtain the weak convergence of $\sqrt{n}(\hat{\theta}_n-\theta_0)$.

We will now calculate its asymptotic variance. It follows from $\Psi(\theta_0)=0$, lemma 6, and (26) that $\sqrt{n}\Psi_n(\theta_0)$ converges weakly to a tight Gaussian process $\mathcal{W}$ in $\ell^\infty(H_p)$ with

$$
\mathrm{Var}(\mathcal{W}(\mathbf{h}))=h_1^T\sigma_1(\mathbf{h})+h_2^T\sigma_2(\mathbf{h})+h_3^T\sigma_3(\mathbf{h})+\int_0^\tau\sigma_4(\mathbf{h})h_4\,\mathrm{d}\Lambda_{10}+\int_0^\tau\sigma_5(\mathbf{h})h_5\,\mathrm{d}\Lambda_{20}.\tag{30}
$$

It follows from (25) and theorem 3.3.1 of van der Vaart & Wellner (1996) that

$$\begin{aligned}&\sigma_1^T(\mathbf{h})(\sqrt{n}(\hat{\alpha}_{cn}-\alpha_{c0}))+\sigma_2^T(\mathbf{h})(\sqrt{n}(\hat{\beta}_{1n}-\beta_{10}))+\sigma_3^T(\mathbf{h})(\sqrt{n}(\hat{\beta}_{2n}-\beta_{20}))\\&\quad+\int_0^\tau\sigma_4(\mathbf{h})\,\mathrm{d}(\sqrt{n}(\hat{\Lambda}_{1n}-\Lambda_{10}))+\int_0^\tau\sigma_5(\mathbf{h})\,\mathrm{d}(\sqrt{n}(\hat{\Lambda}_{2n}-\Lambda_{20}))\\&=-\dot{\Psi}_{\theta_0}(\sqrt{n}(\hat{\theta}_n-(\theta_0))(\mathbf{h})\\&=\sqrt{n}(\Psi_n(\theta_0)-\Psi(\theta_0))(\mathbf{h})+o_{p^*}(1).\end{aligned}\tag{31}$$

Setting $\mathbf{g}=\sigma(\mathbf{h})$ in (31) and using (30), we know that

$$\begin{aligned}&g_1^T(\sqrt{n}(\hat{\alpha}_{cn}-\alpha_{c0}))+g_2^T(\sqrt{n}(\hat{\beta}_{1n}-\beta_{10}))+g_3^T(\sqrt{n}(\hat{\beta}_{2n}-\beta_{20}))\\&\quad+\int_0^\tau g_4\,\mathrm{d}(\sqrt{n}(\hat{\Lambda}_{1n}-\Lambda_{10}))+\int_0^\tau g_5\,\mathrm{d}(\sqrt{n}(\hat{\Lambda}_{2n}-\Lambda_{20}))\end{aligned}$$

is asymptotically normal with mean 0 and variance

$$g_1^T\tilde{\sigma}_1(\mathbf{g})+g_2^T\tilde{\sigma}_2(\mathbf{g})+g_3^T\tilde{\sigma}_3(\mathbf{g})+\int_0^\tau g_4\tilde{\sigma}_4(\mathbf{g})\,\mathrm{d}\Lambda_{10}+\int_0^\tau g_5\tilde{\sigma}_5(\mathbf{g})\,\mathrm{d}\Lambda_{20}.\tag{32}$$

Using (32), we get (29) immediately. This completes the proof.

It follows from lemmas 7, 8 and 9 that the conditions of the theorem 3.1, on rate of convergence, in Murphy & van der Vaart (1999) are satisfied. This together with theorem 2′, we get

Theorem 3′ (Rate of Convergence)

$$||\hat{\Lambda}_{1n,\,\zeta_n}-\Lambda_{10}||_V+||\hat{\Lambda}_{2n,\,\zeta_n}-\Lambda_{20}||_V=O_p^*(||\zeta_n-\zeta_0||+n^{-1/2}).$$

5. Profile likelihood theory

In this section, we focus our attention on the estimation of $(\alpha_c,\beta_1,\beta_2)$, and present the efficient score function, the least favorable submodel, and finally the profile likelihood theory.

For $\mathbf{h}\in H_\infty$, let $\tilde{\sigma}_{123}^T(\mathbf{h})=(\tilde{\sigma}_1(\mathbf{h})^T,\tilde{\sigma}_2(\mathbf{h})^T,\tilde{\sigma}_3(\mathbf{h})^T)$. Let e_i be the $D(\equiv 3d+1)$-dimensional row vector with a 1 in the ith component and zeros elsewhere, for every $i=1,\ldots,D$. Define the $D\times D$ matrix Σ by $\Sigma^{-1}=\left(\tilde{\sigma}_{123}(e_1,0,0),\ldots,\tilde{\sigma}_{123}(e_D,0,0)\right)^T$. We note that Σ is positive definite and symmetric.

We define

$$\ell_{123,\,\theta}[(h_1^T,h_2^T,h_3^T)](X_1,\delta_1,Z_1)=\ell_{1,\,\theta}[h_1](X_1,\delta_1,Z_1)+\ell_{2,\,\theta}[h_2](X_1,\delta_1,Z_1)+\ell_{3,\,\theta}[h_3](X_1,\delta_1,Z_1),$$

for $h_1\in\Re^{d+1}$, $h_2\in\Re^d$, and $h_3\in\Re^d$. Viewing (h_1^T,h_2^T,h_3^T) as a D-dimensional row vector, we can consider $\ell_{123,\,\theta}[\cdot](X_1,\delta_1,Z_1)$ a D-dimensional column vector. $\ell_{123,\,\theta}[\cdot](X_1,\delta_1,Z_1)$ will be abbreviated as $\ell_{123,\,\theta}$.

We also define

$$g_4^*=-\Sigma\begin{pmatrix}\tilde{\sigma}_4(e_1,0,0)\\ \vdots\\ \tilde{\sigma}_4(e_D,0,0)\end{pmatrix}\quad\text{and}\quad g_5^*=-\Sigma\begin{pmatrix}\tilde{\sigma}_5(e_1,0,0)\\ \vdots\\ \tilde{\sigma}_5(e_D,0,0)\end{pmatrix}.$$

Let

$$\tilde{\ell}_0=\ell_{123,\,\theta_0}-\ell_{4,\,\theta_0}[g_4^*]-\ell_{5,\,\theta_0}[g_5^*].\tag{33}$$

Using (26), we can prove $e_i\Sigma^{-1}E(\tilde{\ell}_0(\ell_{4,\,\theta_0}[g_4]+\ell_{5,\,\theta_0}[g_5]))=0$, for every g_4 and g_5 in $BV[0,\tau]$. This shows that $\tilde{\ell}_0$ is the efficient score function for the estimation of $(\alpha_c,\beta_1,\beta_2)$. Similarly, we can also obtain $\Sigma=E\tilde{\ell}_0\tilde{\ell}_0^{\,T}$.

Now we introduce the least favourable submodel. For $\gamma \in \Re^{1\times D}$, the space of D-dimensional row vectors, we define

$$\mathbf{\Lambda}_1(\gamma, \theta; t) = \int_0^t (1 + (\zeta - \gamma) g_4^*)(s)\, \mathrm{d}\Lambda_1(s),$$

$$\mathbf{\Lambda}_2(\gamma, \theta; t) = \int_0^t (1 + (\zeta - \gamma) g_5^*)(s)\, \mathrm{d}\Lambda_2(s),$$

where $\theta = (\zeta, \Lambda_1, \Lambda_2) \in \Theta$. Given θ, the path $\gamma \mapsto (\gamma, \mathbf{\Lambda}_1(\gamma, \theta; \cdot), \mathbf{\Lambda}_2(\gamma, \theta; \cdot))$ defines a parametric submodel, referred to as the submodel indexed by θ. Its log-likelihood for the data (X_1, Δ_1, Z_1), denoted by $\ell(\gamma, \theta; X_1, \Delta_1, Z_1) \equiv \ell(\gamma, \zeta, \Lambda_1, \Lambda_2; X_1, \Delta_1, Z_1)$, equals to $\log L_1(\gamma, \mathbf{\Lambda}_1(\gamma, \theta; \cdot), \mathbf{\Lambda}_2(\gamma, \theta; \cdot);$ $X_1, \Delta_1, Z_1)$. We denote by $\dot{\ell}$ and $\ddot{\ell}$ the first and second derivatives of ℓ in γ, respectively.

Thus, the score function at γ of the submodel indexed by θ, denoted by $\dot{\ell}(\gamma, \theta)$, is equal to $\ell_{123, (\gamma, \mathbf{\Lambda}_1(\gamma, \theta; \cdot), \mathbf{\Lambda}_2(\gamma, \theta; \cdot))} - \ell_{4, (\gamma, \mathbf{\Lambda}_1(\gamma, \theta; \cdot), \mathbf{\Lambda}_2(\gamma, \theta; \cdot))}[g_4^*] - \ell_{5, (\gamma, \mathbf{\Lambda}_1(\gamma, \theta; \cdot), \mathbf{\Lambda}_2(\gamma, \theta; \cdot))}[g_5^*]$. Hence

$$\dot{\ell}(\zeta_0, \theta_0) = \tilde{\ell}_0, \tag{34}$$

which says the score function at $\gamma = \zeta_0$ of the submodel indexed by $\theta_0 = (\zeta_0, \Lambda_{10}, \Lambda_{20})$, is equal to the efficient score function $\tilde{\ell}_0$. Denote by $I_0(\gamma, \theta)$ the Fisher information at γ of the submodel θ. Then

$$I_0(\zeta_0, \theta_0) = -E\ddot{\ell}(\zeta_0, \theta_0) = E\dot{\ell}(\zeta_0, \theta_0)\dot{\ell}^T(\zeta_0, \theta_0) = E\tilde{\ell}_0 \tilde{\ell}_0^T = \Sigma. \tag{35}$$

Lemma 12 below is needed in establishing the profile likelihood theory and we omit its proof because it is technical and involves mainly arguments already used in proving lemmas 2, 3 and 7.

Lemma 12

The functions $(\gamma, \theta) \mapsto \dot{\ell}(\gamma, \theta)(X_1, \Delta_1, Z_1)$ *and* $(\gamma, \theta) \mapsto \ddot{\ell}(\gamma, \theta)(X_1, \Delta_1, Z_1)$ *are continuous at* $(\zeta_0, \zeta_0, \Lambda_1, \Lambda_2)$ *for almost every* (X_1, Δ_1, Z_1)*, relative to the probability specified by* θ_0*. There exists a neighbourhood* V *of* (ζ_0, θ_0) *such that* $\{\dot{\ell}(\gamma, \zeta, \Lambda_1, \Lambda_2) | (\gamma, \zeta, \Lambda_1, \Lambda_2) \in V\}$ *is a uniformly bounded Donsker class, and* $\{\ddot{\ell}(\gamma, \zeta, \Lambda_1, \Lambda_2) | (\gamma, \zeta, \Lambda_1, \Lambda_2) \in V\}$ *is a uniformly bounded Glivenko–Cantelli class.*

Lemma 13 (No bias)

For every random sequence $\tilde{\zeta}_n$ *converging to* ζ_0 *in probability,*

$$E_{\theta_0} \dot{\ell}(\zeta_0, \tilde{\zeta}_n, \hat{\Lambda}_{1n, \tilde{\zeta}_n}, \hat{\Lambda}_{2n, \tilde{\zeta}_n}) = o_p(||\tilde{\zeta}_n - \zeta_0|| + n^{-1/2}).$$

Proof. Since $\dot{\ell}(\zeta, \zeta, \Lambda_1, \Lambda_2)$ is a score function for the submodel indexed by $\theta = (\zeta, \Lambda_1, \Lambda_2)$, we know $E_{(\zeta, \Lambda_1, \Lambda_2)} \dot{\ell}(\zeta, \zeta, \Lambda_1, \Lambda_2) = 0$ for every $(\zeta, \Lambda_1, \Lambda_2)$. Differentiating this identity relative to ζ yields

$$E_{(\zeta, \Lambda_1, \Lambda_2)} \ell_{123, \theta} \dot{\ell}^T(\zeta, \zeta, \Lambda_1, \Lambda_2) + E_{(\zeta, \Lambda_1, \Lambda_2)} \ddot{\ell}(\zeta, \zeta, \Lambda_1, \Lambda_2) + \left.\frac{\partial}{\partial v}\right|_{v=\zeta} E_{(\zeta, \Lambda_1, \Lambda_2)} \dot{\ell}(\zeta, v, \Lambda_1, \Lambda_2) = 0;$$

evaluating this at $(\zeta, \Lambda_1, \Lambda_2) = (\zeta_0, \Lambda_{10}, \Lambda_{20})$ and using the fact that $\dot{\ell}(\zeta_0, \theta_0)$ equals to the efficient score give

$$
\begin{aligned}
&-\frac{\partial}{\partial v}\bigg|_{v=\zeta_0} E_{\theta_0}\dot{\ell}(\zeta_0, v, \Lambda_{10}, \Lambda_{20}) \\
&= E_{\theta_0}\ell_{123}(\zeta_0, \Lambda_{10}, \Lambda_{20})\dot{\ell}^T(\zeta_0, \zeta_0, \Lambda_{10}, \Lambda_{20}) + E_{\theta_0}\ddot{\ell}(\zeta_0, \zeta_0, \Lambda_{10}, \Lambda_{20}) \\
&= 0.
\end{aligned}
$$

Thus

$$
\begin{aligned}
&E_{\theta_0}\left(\dot{\ell}(\zeta_0, \zeta, \Lambda_1, \Lambda_2) - \dot{\ell}(\zeta_0, \zeta_0, \Lambda_1, \Lambda_2)\right) \\
&= E_{\theta_0}\left(\frac{\partial}{\partial v}\bigg|_{v=\zeta_*}\dot{\ell}(\zeta_0, v, \Lambda_1, \Lambda_2) - \frac{\partial}{\partial v}\bigg|_{v=\zeta_0}\dot{\ell}(\zeta_0, v, \Lambda_{10}, \Lambda_{20})\right)(\zeta - \zeta_0)^T,
\end{aligned}
$$

for an intermediate point ζ_* between ζ and ζ_0. This implies

$$
E_{\theta_0}\left(\dot{\ell}(\zeta_0, \zeta, \Lambda_1, \Lambda_2) - \dot{\ell}(\zeta_0, \zeta_0, \Lambda_1, \Lambda_2)\right) = o_p(||\zeta - \zeta_0||). \tag{36}
$$

Since $E_{(\zeta_0, \Lambda_1, \Lambda_2)}\dot{\ell}(\zeta_0, \zeta_0, \Lambda_1, \Lambda_2) = 0$, we know

$$
\begin{aligned}
E_{\theta_0}\dot{\ell}(\zeta_0, \zeta_0, \Lambda_1, \Lambda_2) = {} & (E_{\theta_0} - E_{(\zeta_0, \Lambda_1, \Lambda_2)})(\dot{\ell}(\zeta_0, \zeta_0, \Lambda_1, \Lambda_2) - \dot{\ell}(\zeta_0, \theta_0)) \\
&+ (E_{\theta_0} - E_{(\zeta_0, \Lambda_1, \Lambda_2)})\dot{\ell}(\zeta_0, \theta_0),
\end{aligned} \tag{37}
$$

which is bounded by a multiple of $||\Lambda_1 - \Lambda_{10}||_V + ||\Lambda_2 - \Lambda_{20}||_V$.

It follows from (36), (37) and theorem 3′ that the proof of this lemma is complete.

Let the profile likelihood for ζ be denoted by $pL_n(\zeta)$, which is equal to $\sup_{\substack{\Lambda_1 \in \mathcal{L}_* \\ \Lambda_2 \in \mathcal{L}_*}} L_n(\zeta, \Lambda_1, \Lambda_2)$. It follows from theorem 2′, lemmas 12, 13 and (34) that all the conditions in theorem 1 of Murphy & van der Vaart (2000) are satisfied. Thus, we have

Theorem 4

For every random sequence $\tilde{\zeta}_n$ that converges to ζ_0 in probability,

$$
\begin{aligned}
\log pL_n(\tilde{\zeta}_n) - \log pL_n(\zeta_0) = {} & (\tilde{\zeta}_n - \zeta_0)^T \sum_{i=1}^n \tilde{\ell}_0(X_i, \Delta_i, Z_i) - \frac{1}{2}n(\tilde{\zeta}_n - \zeta_0)^T \Sigma(\tilde{\zeta}_n - \zeta_0) \\
&+ o_{p_{\theta_0}}(\sqrt{n}||(\tilde{\zeta}_n - \zeta_0)^T|| + 1)^2.
\end{aligned} \tag{38}
$$

Here Σ appeared in (35).

Using the consistency of $\hat{\zeta}_n$ given in theorem 2, the invertibility of the efficient Fisher information matrix Σ, and the second order expansion of the profile likelihood (38), we obtain the following three theorems immediately from the profile likelihood theory of Murphy & van der Vaart (2000).

Theorem 5

The NPMLE $\hat{\zeta}_n$ *is asymptotically normal and asymptotically efficient at* (θ_0, Λ_0)*; that is*

$$
\sqrt{n}(\tilde{\zeta}_n - \zeta_0) = \Sigma^{-1}\sqrt{n}P_n\tilde{\ell}_0^T + o_{p_0}(1) \xrightarrow{d} N(0, \Sigma^{-1}).
$$

Theorem 6

Under the null hypothesis $\boldsymbol{H}_0 : \zeta = \zeta_0$, the profile likelihood ratio statistic

$$
lrt_n(\theta_0) \equiv 2\log\frac{pL_n(\hat{\zeta}_n)}{pL_n(\zeta_0)}
$$

is asymptotically chi-squared with D degrees of freedom.

Theorem 7

For all sequences $v_n \xrightarrow{p} v \in \Re^D$ *and* $\rho_n \xrightarrow{p} 0$ *such that* $(\sqrt{n}\rho_n)^{-1} = O_p(1)$,

$$-2\frac{\log pL_n(\hat{\zeta}_n + \rho_n v_n) - \log pL_n(\hat{\zeta}_n)}{n\rho_n^2} \xrightarrow{p} v^T \Sigma v. \tag{39}$$

Using (39), we can show that

$$-\Big[\log pL_n(\hat{\zeta}_n + \rho_n e_i + \rho_n e_j) - \log pL_n(\hat{\zeta}_n + \rho_n e_i)$$
$$- \log pL_n(\hat{\zeta}_n + \rho_n e_j) + \log pL_n(\hat{\zeta}_n)\Big] / (n\rho_n^2)$$

converges in probability to the (i,j) entry of Σ.

6. Simulation studies

There are two studies in this section. In the first study, we set $d=1$; $\alpha_{c_0} = (-2, 5)$, $\beta_{10} = 0.5$, $\beta_{20} = -0.5$, $\lambda_{10}(t) = 1/4$ and $\lambda_{20}(t) = 1/5$; the conditional distribution of the censoring variable C_i given $Z_i = z$ is exponential with parameter $50(1-z)$; the distribution of the covariance Z_i is uniform(0, 1).

There are 1000 replicates in this study and each replicate is a random sample with sample size 300. The number of iteration in using the algorithm in Section 2 is set at 300, and the starting values are set as $\alpha_c = (0, 0), \beta_1 = \beta_2 = 0$, and $\Lambda_1(t) = \Lambda_2(t) = t$.

Based on the data from these 1000 replicates, about 52% of the individuals died, 33% of them were cured and 14% were censored. All the computations are done on an ordinary PC; average computing time needed for one replicate is 9 seconds for our method and 15 seconds for the method of Fine (1999).

Table 1 summarizes the results of this simulation study. The second column of Table 1 lists the true values of the parameters. The third, fourth, fifth, and sixth columns report respectively the sample mean, sample standard deviation (SD) and sample mean-squared error (MSE) of the 1,000 estimates, and the average of the 1,000 standard deviations computed by profile likelihood (SD^{prof}). The final column gives the 95% coverage probability based on the normal approximation.

The numbers in the brackets in the third, fourth and fifth columns are the corresponding results obtained using the method in Fine (1999). Table 1 indicates clearly that our method works quite nicely and the method in Fine (1999) seems to fail. Among other things, the averages of the standard deviations obtained by profile likelihood (column 6) are quite close to the sample standard deviations of the 1,000 estimates (column 4).

The only difference between the first study and the second study is that in the second study, the censoring variable C_1 and covariate Z_1 are independent and C_1 has exponential

Table 1. *Simulation study with censoring variable dependent on covariate*

Parameter	True value	mean	SD	MSE	SD^{prof}	CI(%)
α_1	−2	−2.0437	0.3343	0.1140	0.3345	95.3
		[−1.0511]	[0.2834]	[0.9807]		
α_2	5	5.1444	0.7251	0.5466	0.7395	95.6
		[2.4802]	[0.5060]	[6.6052]		
β_1	0.5	0.4982	0.4341	0.1885	0.4063	92.9
		[1.0948]	[0.4300]	[0.5386]		
β_2	−0.5	−0.4441	0.5456	0.3008	0.5405	93.8
		[−0.0579]	[0.6055]	[0.5621]		

Numbers in brackets are obtained by the method in Fine (1999), and others are obtained by our method.

Table 2. *Simulation study with censoring variable and covariate being independent*

Parameter	True value	mean	SD	MSE	SD^{prof}	CI(%)
α_1	−2	−2.0317	0.3190	0.1028	0.3280	95.9
		[−2.0500]	[0.3688]	[0.1385]		
α_2	5	5.0955	0.6385	0.4168	0.6654	96.6
		[5.1117]	[0.8072]	[0.6641]		
β_1	0.5	0.5036	0.3822	0.1461	0.3649	94.1
		[0.4962]	[0.4054]	[0.1643]		
β_2	−0.5	−0.4657	0.4868	0.2381	0.4876	93.9
		[−0.4988]	[0.5737]	[0.3291]		

Numbers in brackets are obtained by the method in Fine (1999), and others are obtained by our method.

distribution with parameter 25. The results of the second study is reported in Table 2; entries in Table 2 bear the same meaning as those in Table 1. It is clear from Table 2 that both methods work nicely and our method seems to perform a little better in terms of mean-squared error.

7. Application to Taiwan SARS data

We now illustrate our method by analyzing Taiwan SARS data. To keep the discussion brief, we only consider the covariate age. A more complete survival analysis of Taiwan SARS data will be reported elsewhere. On 5 July 2003, Taiwan was removed from the World Health Organization (WHO) list of SARS-affected countries. Among the 664 reported probable cases of SARS, 345 were positive for PCR test of SARS-CoV infection or with sero-conversion ELISA test. The following analysis is based on data for these 345 confirmed cases. Among these 345 cases, 73 of them had died as of October 12, 2003; among these 73 deaths, 37 of them have deaths attributed directly to SARS. Readers are referred to Su (2003) for more information about SARS in Taiwan. In this illustration, we treat deaths not directly attributed to SARS as censored cases. Because the time from onset of symptom to admission became shorter when infection-control measures became stricter and also because of no known treatment, we decide to study distribution of onset-to-death and onset-to-discharge, rather than that of admission-to-death and admission-to-discharge. In the implementation, we set the same number of iterations and the same starting values as those in Section 6.

The results are reported in Table 3 and Figures 1, 2 and 3. The second column of Table 3 reports the estimates; the third and fourth column, respectively, report the 95% confidence intervals and the standard deviation (SD) based on normal approximation. Figure 1 is the plot of case fatality rate against age. Since 95% confidence interval for α_2 is (4.7437, 9.5796), we may conclude with confidence that case fatality rate increases with age. Figures 2 and 3 give the age-specific expected onset-to-death and onset-to-discharge, respectively. We can see that onset-to-death is a decreasing function of age, and onset-to-discharge is an increasing function of age. Because the 95% of β_2 is (−2.6491, −0.6144), it seems that age is an important covariate for onset-to-discharge.

The numbers in the brackets are obtained with the method in Fine (1999). Although the results using Fine (1999) and those using our method are not markedly different, we still prefer our method because it is a more general and systematic approach.

8. Concluding remarks

We have provided a profile likelihood theory and an efficient computational methods for a semiparametric mixture model for competing-risks data, in which proportional hazards

Table 3. *Parameter estimation for Taiwan SARS data*

Parameter	Estimate	95% confidence interval	SD
α_1	−5.3554 [−5.1539]	(−6.7901,−3.9207)	0.7320
α_2	7.1617 [6.3673]	(4.7437, 9.5796)	1.2337
β_1	0.9988 [0.8488]	(−1.7824, 3.7800)	1.4190
β_2	−1.6818 [−1.0989]	(−2.6491, −0.7144)	0.4936

Numbers in brackets are obtained by the method in Fine (1999), and others are obtained by our method.

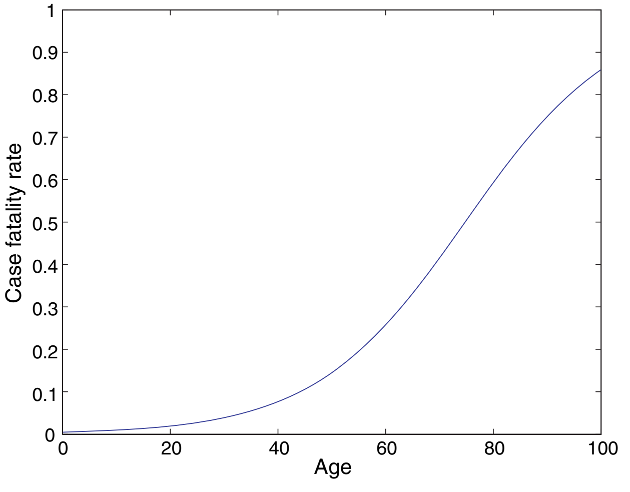

Fig. 1. Age-specific case fatality rate.

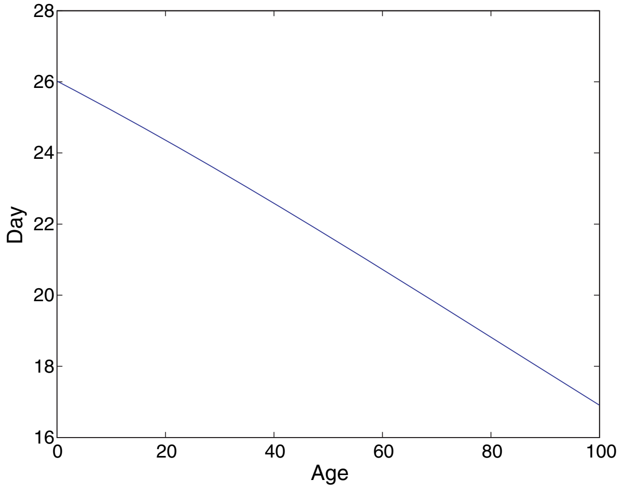

Fig. 2. Age-specific expected onset-to-death.

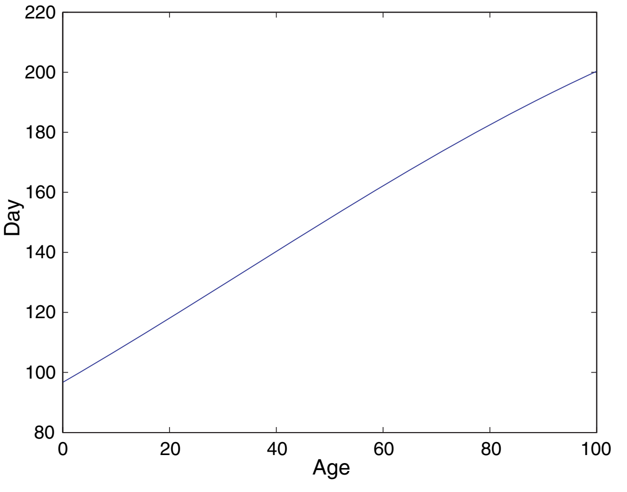

Fig. 3. Age-specific expected onset-to-discharge.

models are specified for failure time models conditional on cause and a multinomial model for the marginal distribution of cause conditional on covariates. We note that our model allows data to be right censored and the censoring variable may depend on the covariates. We have also successfully illustrated this theory in simulation studies and in the analysis of Taiwan SARS data. Our approach can be used to estimate quantities like covariate specific fatality rate and covariate-specific expected time from onset to death. For future investigation, we would like to extend the present work to semiparametric mixture models in which the failure time models conditional on cause are replaced by other survival models for right-censored data.

Acknowledgements

We thank Professor Bickel for suggesting the possibility of adapting the computational strategy in Chang *et al.* (2006) for this paper, and thank CDC Taiwan for making available the data to illustrate the method. We are grateful to the associate editor and referee for their critical comments and suggestions that lead to improvement of this paper. This research is partially supported by Taiwan National Science Council Grant 95-2119-M-032-001.

References

Andersen, P. K. (2002). Competing risks as a multi-state model. *Stat. Methods Med. Res.* **11**, 203–215.

Betensky, R. A. & Schoenfeld, D. A. (2001). Nonparametric estimation in a cure model with random cure times. *Biometrics* **57**, 282–286.

Chang, I. S., Hsiung, 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.

Chang, I. S., Wen, C. C., Wu, Y. J. & Yang, C. C. (2006). A fast algorithm for the nonparameteric maximum likelihood estimate in the Cox-gene model. Accepted by *Statist. Sinica*.

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

Donnelly, C. A., Ghanl, A. C., Leung, G. M., Hedley, A. J., Fraser, C., Rifey, S., Abu-Radded, L. J., Ho, L. M., Thach, T. Q., Chau, P., Chan, K. P., Lam, T. H., Tse, L. Y., Tsang, T., Llu, S. H., Kong, J. H. B., Lau, E. M. C., Ferguson, N. M. & Anderson, R. M. (2003). Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet* **361**, 1761–1766.

Farewell, V. T. (1977). A model for a binary variable with time-censored observations. *Biometrika* **64**, 43–46.

Fine, J. P. (1999). Analyzing competing risks data with transformation models. *J. Roy. Stat. Soc. Ser. B Stat. Methodol.* **61**, 817–830.

Fine, J. P. (2001). Regression modeling of competing crude failure probabilities. *Biostatistics* **2**, 85–97.

Fine, J. P. & Gray, R. (1999). A proportional hazards model for the subdistribution of a competing risk. *J. Amer. Statist. Assoc.* **94**, 496–509.

Kosorok, M., Lee, B. & Fine, J. (2004). Robust inference for proportional hazards univariate frailty regression models. *Ann. Statist*. **32**, 1448–1491.

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

Kuk, A. Y. C. & Chen, C. (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. *J. Roy. Statist. Soc. Ser. C* **34**, 201–211.

Li, C.-S., Taylor, M. G. & Judy, P. S. (2001). Identifiability of cure models. *Statist. Probab. Lett.* **54**, 389–395.

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

Murphy, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22**, 712–731.

Murphy, S. A., Rossini, A. J. & van der Vaart, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *J. Amer. Statist. Assoc.* **92**, 968–976.

Murphy, S. A. & van der Vaart, A. W. (1999). Observed information in semi-parametric models. *Bernoulli* **5**, 381–412.

Murphy, S. A. & van der Vaart, A. W. (2000). On profile likelihood. *J. Amer. Statist. Assoc.* **95**, 449–465.

Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *Ann. Statist.* **26**, 183–214.

Peng, Y. (2003). Fitting semiparametric cure models. *Comput. Statist. Data Anal.* **41**, 481–490.

Peng, Y. & Dear, K. B. G. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics* **56**, 237–243.

Robins, J. M. & Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS epidemiology-methodological issues* (eds N. Jewell, K. Dietz & V. Farewell), 21–33. Birkhauser, Boston.

Rudin, W. (1973). *Functional analysis*. McGraw-Hill, New York.

Rudin, W. (1976). *Principles of mathematical analysis*. 3rd edn. McGraw-Hill, New York.

Su, I. J. (2003). *Memoir of severe acute respiratory syndrome control in Taiwan*. Center for Disease Control, Department of Health, Taiwan.

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.

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

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

Received July 2004, in final form April 2007

I-Shou Chang, Institute of Cancer Research and Division of Biostatistics and Bioinformatics, National Health Research Institutes, 35, Keyan Road, Zhunan Town, Miaoli County 350, Taiwan.
E-mail: ischang@nhri.org.tw

Appendix

Proposition (Identifiability)
Assumption 2 *implies assumption* 1.

Proof. By considering (4) for X_1 near t^* and $\Delta_1 = 1$, we can show the existence of $\lim_{t \to t^*} \frac{d\Lambda_1}{d\Lambda_{10}}(t)$; denote it by y_1^*. Similarly, let $y_2^* = \lim_{t \to t^*} \frac{d\Lambda_2}{d\Lambda_{20}}(t)$. By considering (4) for $[C_1 \geq \tau, \Delta_1 = 1]$, we can show that $(\alpha_c, \beta_1) = (\alpha_{c0}, \beta_{10})$ implies $\Lambda_1 = \Lambda_{10}$ on $[0, \tau]$. Similarly, we know $(\alpha_c, \beta_2) = (\alpha_{c0}, \beta_{20})$ implies $\Lambda_2 = \Lambda_{20}$ on $[0, \tau]$.

Considering now X_1 near t^* in (4), we get

$$\begin{aligned}
&\{\alpha(Z_1) y_1^* \exp(\beta_1^T Z_1 - \Lambda_1(t^*) e^{\beta_1^T Z_1})\}^{[\triangle_1 = 1]} \\
&\quad \times \{[1 - \alpha(Z_1)] y_2^* \exp(\beta_2^T Z_1 - \Lambda_2(t^*) e^{\beta_2^T Z_1})\}^{[\triangle_1 = 2]} \\
&\quad \times \{\alpha(Z_1) \exp(-\Lambda_1(t^*) e^{\beta_1^T Z_1}) + [1 - \alpha(Z_1)] \exp(-\Lambda_2(t^*) e^{\beta_2^T Z_1})\}^{[\triangle_1 = 3]} \\
&= \{\alpha_0(Z_1) \exp(\beta_{10}^T Z_1 - \Lambda_{10}(t^*) e^{\beta_{10}^T Z_1})\}^{[\triangle_1 = 1]} \\
&\quad \times \{[1 - \alpha_0(Z_1)] \exp(\beta_{20}^T Z_1 - \Lambda_{20}(t^*) e^{\beta_{20}^T Z_1})\}^{[\triangle_1 = 2]} \\
&\quad \times \{\alpha_0(Z_1) \exp(-\Lambda_{10}(t^*) e^{\beta_{10}^T Z_1}) + [1 - \alpha_0(Z_1)] \exp(-\Lambda_{20}(t^*) e^{\beta_{20}^T Z_1})\}^{[\triangle_1 = 3]}.
\end{aligned} \tag{40}$$

It follows from (40) and assumption 2 that $\alpha_c = \alpha_{c0}, \beta_1 = \beta_{10}$ and $\beta_2 = \beta_{20}$. This completes the proof.