Statistica Sinica **20** (2010), 871-910

A GENERAL ASYMPTOTIC THEORY FOR MAXIMUM LIKELIHOOD ESTIMATION IN SEMIPARAMETRIC REGRESSION MODELS WITH CENSORED DATA

Donglin Zeng and D. Y. Lin

University of North Carolina at Chapel Hill

Abstract: We establish a general asymptotic theory for nonparametric maximum likelihood estimation in semiparametric regression models with right censored data. We identify a set of regularity conditions under which the nonparametric maximum likelihood estimators are consistent, asymptotically normal, and asymptotically efficient with a covariance matrix that can be consistently estimated by the inverse information matrix or the profile likelihood method. The general theory allows one to obtain the desired asymptotic properties of the nonparametric maximum likelihood estimators for any specific problem by verifying a set of conditions rather than by proving technical results from first principles. We demonstrate the usefulness of this powerful theory through a variety of examples.

Key words and phrases: Counting process, empirical process, multivariate failure times, nonparametric likelihood, profile likelihood, survival data.

1. Introduction

Semiparametric regression models are highly useful in investigating the effects of covariates on potentially censored responses (e.g., failure times and repeated measures) in longitudinal studies. It is desirable to analyze such models by the nonparametric maximum likelihood approach, which generally yields consistent, asymptotically normal, and asymptotically efficient estimators. It is technically difficult to prove the asymptotic properties of the nonparametric maximum likelihood estimators (NPMLEs). Thus far, rigorous proofs exist only in some special cases.

In this paper, we develop a general asymptotic theory for the NPMLEs with right censored data. The theory is very encompassing in that it pertains to a generic form of likelihood rather than specific models. We prove that, under a set of mild regularity conditions, the NPMLEs are consistent, asymptotically normal, and asymptotically efficient with a limiting covariance matrix that can be consistently estimated by the inverse information matrix or the profile likelihood method.

This paper is the technical companion to Zeng and Lin (2007), in which several classes of models were proposed to unify and extend existing semiparametric regression models. The likelihoods for those models can all be written in the general form considered in this paper. For each class of models in Zeng and Lin (2007), we identify a set of conditions under which the regularity conditions for the general theory hold so that desired asymptotic properties are ensured.

2. Some Semiparametric Models

We describe briefly the three kinds of models considered in Zeng and Lin (2007). We assume that the censoring mechanism satisfies coarsening at random (Heitjan and Rubin (1991)).

2.1. Transformation models for counting processes

Let $N^*(t)$ record the number of events that the subject has experienced by time t, and let $Z(\cdot)$ denote the corresponding covariate processes. Zeng and Lin (2007) proposed the following class of transformation models for the cumulative intensity function of $N^*(t)$

$$\Lambda(t|Z) = G\left[\left\{1 + \int_0^t R^*(s) e^{\beta^T Z(s)} d\Lambda(s)\right\}^{e^{\gamma^T \widetilde{Z}}}\right] - G(1),$$

where G is a continuously differentiable and strictly increasing function with $G'(1) > 0$ and $G(\infty) = \infty$, $R^*(\cdot)$ is an indicator process, $\widetilde{Z}$ is a subset of Z, β and γ are regression parameters, and $\Lambda(\cdot)$ is an unspecified increasing function. The data consist of $\{N_i(t), R_i(t), Z_i(t); t \in [0, \tau]\}$ $(i = 1, \ldots, n)$, where $R_i(t) = I(C_i \geq t) R_i^*(t)$, $N_i(t) = N_i^*(t \wedge C_i)$, C_i is the censoring time, and τ is a finite constant. The likelihood is

$$\prod_{i=1}^n \prod_{t \leq \tau} \{R_i(t) d\Lambda(t|Z_i)\}^{dN_i(t)} \exp\left\{-\int_0^\tau R_i(t) d\Lambda(t|Z_i)\right\},$$

where $dN_i(t) = N_i(t) - N_i(t-)$.

2.2. Transformation models with random effects for dependent failure times

For $i = 1, \ldots, n$, $k = 1, \ldots, K$ and $l = 1, \ldots, n_{ik}$, let $N_{ikl}^*(\cdot)$ denote the number of the kth type of event experienced by the lth individual in the ith cluster, and $Z_{ikl}(\cdot)$ the corresponding covariate processes. Zeng and Lin (2007) assumed that the cumulative intensity for $N_{ikl}^*(t)$ takes the form

$$\Lambda_k(t|Z_{ikl}; b_i) = G_k\left\{\int_0^t R_{ikl}^*(s) e^{\beta^T Z_{ikl}(s) + b_i^T \widetilde{Z}_{ikl}(s)} d\Lambda_k(s)\right\},$$

where G_k, Λ_k, and R^*_{ikl} are analogous to G, Λ, and R^* of Section 2.1, $\widetilde{Z}_{ikl}$ is a subset of Z_{ikl} plus the unit component, and b_i is a vector of random effects with density $f(b;\gamma)$. Let C_{ikl}, N_{ikl}, and R_{ikl} be defined analogously to C_i, N_i, and R_i of Section 2.1. The likelihood is

$$\prod_{i=1}^{n}\int_b \prod_{k=1}^{K}\prod_{l=1}^{n_{ik}}\prod_{t\le\tau}\Big[R_{ikl}(t)e^{\beta^T Z_{ikl}(t)+b^T\widetilde{Z}_{ikl}(t)}d\Lambda_k(t)$$
$$\times G'_k\Big\{\int_0^t R_{ikl}(s)e^{\beta^T Z_{ikl}(s)+b^T\widetilde{Z}_{ikl}(s)}d\Lambda_k(s)\Big\}\Big]^{dN_{ikl}(t)}$$
$$\times\exp\Big[-G_k\Big\{\int_0^\tau R_{ikl}(t)e^{\beta^T Z_{ikl}(t)+b^T\widetilde{Z}_{ikl}(t)}d\Lambda_k(t)\Big\}\Big]f(b;\gamma)db.$$

2.3. Joint models for repeated measures and failure times

For $i=1,\ldots,n$ and $j=1,\ldots,n_i$, let Y_{ij} be the response variable at time t_{ij} for the ith subject, and X_{ij} the corresponding covariates. We assume that $(Y_{i1},\ldots,Y_{in_i})$ follows a generalized linear mixed model with density $f_y(y|X_{ij};b_i)$, where b_i is a set of random effects with density $f(b;\gamma)$. We define N^*_i and Z_i as in Section 2.1, and assume that

$$\Lambda(t|Z_i;b_i)=G\Big\{\int_0^t R^*_i(s)e^{\beta^T Z_i(s)+(\psi\circ b_i)^T\widetilde{Z}_i(s)}d\Lambda(s)\Big\},$$

where $\widetilde{Z}_i$ is a subset of Z_i plus the unit component, ψ is a vector of unknown constants, and $v_1\circ v_2$ is the component-wise product of two vectors v_1 and v_2. The likelihood is

$$\prod_{i=1}^{n}\int_b\prod_{t\le\tau}\Big\{R_i(t)d\Lambda(t|Z_i;b)\Big\}^{dN_i(t)}\exp\Big\{-\int_0^\tau R_i(t)d\Lambda(t|Z_i;b)\Big\}$$
$$\times\prod_{j=1}^{n_i}f_y(Y_{ij}|X_{ij};b)f(b;\gamma)db.$$

For continuous measures, Zeng and Lin (2007) proposed the semiparametric linear mixed model

$$\widetilde{H}(Y_{ij})=\alpha^T X_{ij}+b_i^T\widetilde{X}_{ij}+\epsilon_{ij},$$

where $\widetilde{H}$ is an unknown increasing function with $\widetilde{H}(-\infty)=-\infty$, $\widetilde{H}(\infty)=\infty$, and $\widetilde{H}(0)=0$, α is a set of regression parameters, $\widetilde{X}_{ij}$ is typically a subset of X_{ij}, and ϵ_{ij} $(i=1,\ldots,n;j=1,\ldots,n_{ij})$ are independent with density f_ϵ. Write $\widetilde{\Lambda}(y)=e^{\widetilde{H}(y)}$. The likelihood is

$$\prod_{i=1}^{n}\int_b\prod_{t\le\tau}\Big\{R_i(t)d\Lambda(t|Z_i;b)\Big\}^{dN_i(t)}\exp\Big\{-\int_0^\tau R_i(t)d\Lambda(t|Z_i;b)\Big\}$$

$$\times \prod_{j=1}^{n_i} f_\epsilon\Big(\log(\widetilde{\Lambda}(Y_{ij})) - \alpha^T X_{ij} - b_i^T \widetilde{X}_{ij}\Big)\left\{\frac{d\log\widetilde{\Lambda}(Y_{ij})}{dy}\right\} f(b;\gamma)db.$$

3. Nonparametric Maximum Likelihood Estimation

All the likelihood functions given in Section 2 can be expressed as

$$\prod_{i=1}^{n}\prod_{k=1}^{K}\prod_{l=1}^{n_{ik}}\prod_{t\le\tau} \lambda_k(t)^{R_{ikl}(t)dN^*_{ikl}(t)}\Psi(\mathcal{O}_i;\theta,\mathcal{A}),$$

where $\lambda_k(t) = \Lambda'_k(t)$, θ is a d-vector of regression parameters and variance components, $\mathcal{A} = (\Lambda_1, \ldots, \Lambda_K)$, $\mathcal{O}_i$ pertains to the observation on the ith cluster, and Ψ is a functional of $\mathcal{O}_i$, θ, and $\mathcal{A}$. For nonparametric maximum likelihood estimation, we allow $\mathcal{A}$ to be discontinuous with jumps at the observed failure times and maximize the modified likelihood function

$$\prod_{i=1}^{n}\prod_{k=1}^{K}\prod_{l=1}^{n_{ik}}\prod_{t\le\tau} \Lambda_k\{t\}^{R_{ikl}(t)dN^*_{ikl}(t)}\Psi(\mathcal{O}_i;\theta,\mathcal{A}),$$

where $\Lambda_k\{t\}$ denotes the jump size of the monotone function Λ_k at t. Equivalently, we maximize the logarithm of the above function

$$\mathcal{L}_n(\theta,\mathcal{A}) = \sum_{i=1}^{n}\left[\sum_{k=1}^{K}\sum_{l=1}^{n_{ik}}\int_0^\tau R_{ikl}(t)\log\Lambda_k\{t\}dN^*_{ikl}(t) + \log\Psi(\mathcal{O}_i;\theta,\mathcal{A})\right]. \quad (3.1)$$

We wish to establish an asymptotic theory for the resulting NPMLEs $\widehat{\theta}$ and $\widehat{\mathcal{A}}$.

4. Regularity Conditions

We impose the following conditions on the model and data structures.

(C1) The true value θ_0 lies in the interior of a compact set Θ, and the true functions Λ_{0k} are continuously differentiable in $[0,\tau]$ with $\Lambda'_{0k}(t) > 0$, $k = 1,\ldots,K$.

(C2) With probability one, $P(\inf_{s\in[0,t]} R_{ik\cdot}(s) \ge 1 | Z_{ikl},\ l = 1,\ldots,n_{ik}) > \delta_0 > 0$ for all $t \in [0,\tau]$, where $R_{ik\cdot}(t) = \sum_{l=1}^{n_{ik}} R_{ikl}(t)$.

(C3) There exist a constant $c_1 > 0$ and a random variable $r_1(\mathcal{O}_i) > 0$ such that $E[\log r_1(\mathcal{O}_i)] < \infty$ and, for any $\theta \in \Theta$ and any finite $\Lambda_1,\ldots,\Lambda_K$,

$$\Psi(\mathcal{O}_i;\theta,\mathcal{A}) \le r_1(\mathcal{O}_i)\prod_{k=1}^{K}\prod_{t\le\tau}\left\{1 + \int_0^t R_{ik\cdot}(t)d\Lambda_k(t)\right\}^{-dN^*_{ik\cdot}(t)}$$

$$\times\left\{1+\int_0^\tau R_{ik\cdot}(t)d\Lambda_k(t)\right\}^{-c_1}$$

almost surely, where $N^*_{ik\cdot}(t)=\sum_{l=1}^{n_{ik}} N^*_{ikl}(t)$. In addition, for any constant c_2,

$$\inf\left\{\Psi(\mathcal{O}_i;\theta,\mathcal{A}):\|\Lambda_1\|_{V[0,\tau]}\le c_2,\ldots,\|\Lambda_K\|_{V[0,\tau]}\le c_2,\theta\in\Theta\right\}>r_2(\mathcal{O}_i)>0,$$

where $\|h\|_{V[0,\tau]}$ is the total variation of $h(\cdot)$ in $[0,\tau]$, and $r_2(\mathcal{O}_i)$, which may depend on c_2, is a finite random variable with $E[|\log r_2(\mathcal{O}_i)|]<\infty$.

We require certain smoothness of Ψ. Let $\dot\Psi_\theta$ denote the derivative of $\Psi(\mathcal{O}_i;\theta,\mathcal{A})$ with respect to θ, and let $\dot\Psi_k[H_k]$ denote the derivative of $\Psi(\mathcal{O}_i;\theta,\mathcal{A})$ along the path $(\Lambda_k+\epsilon H_k)$, where H_k belongs to the set of functions in which $\Lambda_k+\epsilon H_k$ is increasing with bounded total variation.

(C4) For any $(\theta^{(1)},\theta^{(2)})\in\Theta$, and $(\Lambda_1^{(1)},\Lambda_1^{(2)})$, …, $(\Lambda_K^{(1)},\Lambda_K^{(2)})$, $(H_1^{(1)},H_1^{(2)})$, …, $(H_K^{(1)},H_K^{(2)})$ with uniformly bounded total variations, there exist a random variable $\mathcal{F}(\mathcal{O}_i)\in L_4(P)$ and K stochastic processes $\mu_{ik}(t;\mathcal{O}_i)\in L_6(P)$, $k=1,\ldots,K$, such that

$$\begin{aligned}
&\left|\Psi\Big(\mathcal{O}_i;\theta^{(1)},\mathcal{A}^{(1)}\Big)-\Psi\Big(\mathcal{O}_i;\theta^{(2)},\mathcal{A}^{(2)}\Big)\right|+\left|\dot\Psi_\theta\Big(\mathcal{O}_i;\theta^{(1)},\mathcal{A}^{(1)}\Big)-\dot\Psi_\theta\Big(\mathcal{O}_i;\theta^{(2)},\mathcal{A}^{(2)}\Big)\right|\\
&\quad+\sum_{k=1}^K\left|\dot\Psi_k\Big(\mathcal{O}_i;\theta^{(1)},\mathcal{A}^{(1)}\Big)\Big[H_k^{(1)}\Big]-\dot\Psi_k\Big(\mathcal{O}_i;\theta^{(2)},\mathcal{A}^{(2)}\Big)\Big[H_k^{(2)}\Big]\right|\\
&\quad+\sum_{k=1}^K\left|\frac{\dot\Psi_k(\mathcal{O}_i;\theta^{(1)},\mathcal{A}^{(1)})[H_k^{(1)}]}{\Psi(\mathcal{O}_i;\theta^{(1)},\mathcal{A}^{(1)})}-\frac{\dot\Psi_k(\mathcal{O}_i;\theta^{(2)},\mathcal{A}^{(2)})[H_k^{(2)}]}{\Psi(\mathcal{O}_i;\theta^{(2)},\mathcal{A}^{(2)})}\right|\\
&\le\mathcal{F}(\mathcal{O}_i)\Bigg[\left|\theta^{(1)}-\theta^{(2)}\right|+\sum_{k=1}^K\Bigg\{\int_0^\tau\left|\Lambda_k^{(1)}(s)-\Lambda_k^{(2)}(s)\right|d\mu_{ik}(s;\mathcal{O}_i)\\
&\quad+\int_0^\tau\left|H_k^{(1)}(s)-H_k^{(2)}(s)\right|d\mu_{ik}(s;\mathcal{O}_i)\Bigg\}\Bigg].
\end{aligned}$$

In addition, $\mu_{ik}(s;\mathcal{O}_i)$ is non-decreasing, and $E[\mathcal{F}(\mathcal{O}_i)\mu_{ik}(s;\mathcal{O}_i)]$ is left-continuous with uniformly bounded left- and right-derivatives for any $s\in[0,\tau]$. Here, the right-derivative for a function $f(x)$ is defined as $\lim_{h\to0+}(f(x+h)-f(x+))/h$.

The following condition ensures identifiability of parameters.

(C5) (First Identifiability Condition) If

$$\left[\prod_{k=1}^K\prod_{l=1}^{n_{ik}}\prod_{t\le\tau}\lambda_k^*(t)^{R_{ikl}(t)dN^*_{ikl}(t)}\right]\Psi(\mathcal{O}_i;\theta^*,\mathcal{A}^*)$$

$$= \left[\prod_{k=1}^{K} \prod_{l=1}^{n_{ik}} \prod_{t \le \tau} \lambda_{0k}(t)^{R_{ikl}(t) dN_{ikl}^*(t)} \right] \Psi(\mathcal{O}_i; \theta_0, \mathcal{A}_0)$$

almost surely, then $\theta^* = \theta_0$ and $\Lambda_k^*(t) = \Lambda_{0k}(t)$ for $t \in [0, \tau]$, $k = 1, \ldots, K$.

The next assumption is more technical and will be used in proving the weak convergence of the NPMLEs. For any fixed $(\theta, \mathcal{A})$ in a small neighborhood of $(\theta_0, \mathcal{A}_0)$ in $R^d \times \{BV[0, \tau]\}^K$, where $BV[0, \tau]$ denotes the space of functions with bounded total variations in $[0, \tau]$, (C4) implies that the linear functional

$$H_k \mapsto E\left[\frac{\dot{\Psi}_k(\mathcal{O}_i; \theta, \mathcal{A})[H_k]}{\Psi(\mathcal{O}_i; \theta, \mathcal{A})} \right]$$

is continuous from $BV[0, \tau]$ to R. Thus, there exists a bounded function $\eta_{0k}(s; \theta, \mathcal{A})$ such that

$$E\left[\frac{\dot{\Psi}_k(\mathcal{O}_i; \theta, \mathcal{A})[H_k]}{\Psi(\mathcal{O}_i; \theta, \mathcal{A})} \right] = \int_0^\tau \eta_{0k}(s; \theta, \mathcal{A}) dH_k(s).$$

(C6) There exist functions $\zeta_{0k}(s; \theta_0, \mathcal{A}_0) \in BV[0, \tau]$, $k = 1, \ldots, K$, and a matrix $\zeta_{0\theta}(\theta_0, \mathcal{A}_0)$ such that

$$\begin{aligned} &\Bigg| E\left[\frac{\dot{\Psi}_\theta(\mathcal{O}_i; \theta, \mathcal{A})}{\Psi(\mathcal{O}_i; \theta, \mathcal{A})} - \frac{\dot{\Psi}_\theta(\mathcal{O}_i; \theta_0, \mathcal{A}_0)}{\Psi(\mathcal{O}_i; \theta_0, \mathcal{A}_0)} \right] - \zeta_{0\theta}(\theta_0, \mathcal{A}_0)(\theta - \theta_0) \\ &\quad - \sum_{k=1}^{K} \int_0^\tau \zeta_{0k}(s; \theta_0, \mathcal{A}_0) d(\Lambda_k - \Lambda_{0k}) \Bigg| \\ &= o\left(|\theta - \theta_0| + \sum_{k=1}^{K} \|\Lambda_k - \Lambda_{0k}\|_{V[0,\tau]} \right). \end{aligned}$$

In addition, for $k = 1, \ldots, K$,

$$\begin{aligned} &\sum_{k=1}^{K} \sup_{s \in [0,\tau]} \Bigg| \Big\{ \eta_{0k}(s; \theta, \mathcal{A}) - \eta_{0k}(s; \theta_0, \mathcal{A}_0) \Big\} - \eta_{0k\theta}(s; \theta_0, \mathcal{A}_0)(\theta - \theta_0) \\ &\quad - \int_0^\tau \sum_{m=1}^{K} \eta_{0km}(s, t; \theta_0, \mathcal{A}_0) d(\Lambda_m - \Lambda_{0m})(t) \Bigg| \\ &= o\left(|\theta - \theta_0| + \sum_{k=1}^{K} \|\Lambda_k - \Lambda_{0k}\|_{V[0,\tau]} \right), \end{aligned}$$

where η_{0km} is a bounded bivariate function and $\eta_{0k\theta}$ is a d-dimensional bounded function. Furthermore, there exists a constant c_3 such that $|\eta_{0km}(s, t_1; \theta_0, \mathcal{A}_0) - \eta_{0km}(s, t_2; \theta_0, \mathcal{A}_0)| \le c_3 |t_1 - t_2|$ for any $s \in [0, \tau]$ and any $t_1, t_2 \in [0, \tau]$.

The final assumption ensures that the Fisher information matrix along any finite-dimensional submodel is non-singular.

(C7) (Second Identifiability Condition) If with probability one,

$$\sum_{k=1}^{K}\sum_{l=1}^{n_{ik}}\int h_k(t)R_{ikl}(t)dN^*_{ikl}(t) \\ +\frac{\dot\Psi_\theta(\mathcal{O}_i;\theta_0,\mathcal{A}_0)^T v+\sum_{k=1}^K\dot\Psi_k(\mathcal{O}_i;\theta_0,\mathcal{A}_0)[\int h_k d\Lambda_{0k}]}{\Psi(\mathcal{O}_i;\theta_0,\mathcal{A}_0)}=0$$

for some constant vector $v\in R^d$ and $h_k\in BV[0,\tau]$, $k=1,\ldots,K$, then $v=0$ and $h_k=0$ for $k=1,\ldots,K$.

Remark 1. (C1)$-$(C2) are standard assumptions in any analysis of censored data. (C3) pertains to the model structure, and (C4) and (C6) essentially impose the smoothness of this structure. Although they appear technical, these conditions are easy to verify in practice. (C5) and (C7) usually require some work to verify, but can be translated to simple conditions in specific cases.

5. Some Useful Lemmas

Lemma 1. *For any constant c, the following classes of functions are P-Donsker:*

$$\mathcal{F}_1=\Big\{\log\Psi(\mathcal{O}_i;\theta,\mathcal{A}):\|\Lambda_k\|_{V[0,\tau]}\le c,\ k=1,\ldots,K,\ \theta\in\Theta\Big\},$$
$$\mathcal{F}_2=\left\{\frac{\dot\Psi_\theta(\mathcal{O}_i;\theta,\mathcal{A})}{\Psi(\mathcal{O}_i;\theta,\mathcal{A})}:\|\Lambda_k\|_{V[0,\tau]}\le c,\ k=1,\ldots,K,\ \theta\in\Theta\right\},$$
$$\mathcal{F}_{3k}=\left\{\frac{\dot\Psi_k(\mathcal{O}_i;\theta,\mathcal{A})[H]}{\Psi(\mathcal{O}_i;\theta,\mathcal{A})}:\|\Lambda_m\|_{V[0,\tau]}\le c,\ m=1,\ldots,K,\ \theta\in\Theta,\ \|H\|_{V[0,\tau]}\le c\right\},\\ k=1,\ldots,K.$$

Proof. We only prove that $\mathcal{F}_{3k}$ is P-Donsker, the proofs for the other two classes are similar. For $k=1,\ldots,K$, we define a measure $\widetilde\mu_k$ in $[0,\tau]$ such that, for any Borel set $A\subset[0,\tau]$,

$$\widetilde\mu_k(A)=\int_0^\tau I(t\in A)E\Big[\mathcal{F}(\mathcal{O}_i)^2\Big(\mu_{ik}(\tau;\mathcal{O}_i)-\mu_{ik}(0;\mathcal{O}_i)\Big)^2d\mu_{ik}(t;\mathcal{O}_i)\Big].$$

Condition (C4) implies that $\widetilde\mu_k([0,\tau])\le\|\mathcal{F}(\mathcal{O}_i)\|_{L_4(P)}\|\mu_{ik}(\tau;\mathcal{O}_i)-\mu_{ik}(0;\mathcal{O}_i)\|_{L_6(P)}$. Thus, $\widetilde\mu_k$ is a finite measure. According to Theorem 2.7.5 of Van der Vaart and Wellner (1996), the bracket covering number for any bounded set in $BV[0,\tau]$ is of order $\exp\{O(1/\epsilon)\}$ in $L_2(\widetilde{\mu}_k)$, $k=1,\ldots,K$. Thus, we can construct $N_\epsilon \equiv (1/\epsilon)^d \times \exp\{O(K/\epsilon)\} \times \exp\{O(1/\epsilon)\}$ brackets for the set of $(\theta, \mathcal{A}, H)$ in $\mathcal{F}_{3k}$, denoted by

$$\left[\theta_p^L, \theta_p^U\right] \times \left[\Lambda_{1p}^L, \Lambda_{1p}^U\right] \times \cdots \times \left[\Lambda_{Kp}^L, \Lambda_{Kp}^U\right] \times \left[H_p^L, H_p^U\right], \quad p = 1, \ldots, N_\epsilon,$$

such that $|\theta_p^U - \theta_p^L| < \epsilon$ and

$$\int \left|\Lambda_{kp}^U - \Lambda_{kp}^L\right|^2 d\widetilde{\mu}_k < \epsilon^2, \quad \int \left|H_p^U - H_p^L\right|^2 d\widetilde{\mu}_k < \epsilon^2, \quad k = 1, \ldots, K.$$

Any $(\theta, \mathcal{A}, H)$ must belong to one of these brackets. Obviously, the bracket functions

$$\frac{\dot{\Psi}_k(\mathcal{O}_i; \theta_p^L, \mathcal{A}_p^L)[H^L]}{\Psi(\mathcal{O}_i; \theta_p^L, \mathcal{A}_p^L)} \pm \mathcal{F}(\mathcal{O}_i)\Bigg\{\left|\theta_p^U - \theta_p^L\right| + \sum_{m=1}^K \int \left|\Lambda_{mp}^U(s) - \Lambda_{mp}^L(s)\right| \\ \times d\mu_{im}(s; \mathcal{O}_i) + \int \left|H_p^U(s) - H_p^L(s)\right| d\mu_{im}(s; \mathcal{O}_i)\Bigg\}, \quad p = 1, \ldots, N_\epsilon,$$

cover all the functions in $\mathcal{F}_{3k}$. Since

$$\begin{aligned}
&\Bigg\| \mathcal{F}(\mathcal{O}_i)\Bigg\{\left|\theta_p^U - \theta_p^L\right| + \sum_{m=1}^K \int \left|\Lambda_{mp}^U(s) - \Lambda_{mp}^L(s)\right| d\mu_{im}(s; \mathcal{O}_i) \\
&\quad + \sum_{m=1}^K \int \left|H_p^U(s) - H_p^L(s)\right| d\mu_{im}(s; \mathcal{O}_i)\Bigg\}\Bigg\|_{L_2(P)} \\
&\quad \le c\Bigg[\left|\theta_p^U - \theta_p^L\right| + \sum_{m=1}^K \Bigg\{E\Bigg(\int \left|\Lambda_{mp}^U(s) - \Lambda_{mp}^L(s)\right| d\widetilde{\mu}_{im} \mathcal{F}(\mathcal{O}_i)\Bigg)^2\Bigg\}^{1/2} \\
&\quad + \sum_{m=1}^K \Bigg\{\int_0^\tau \left|H_p^U(s) - H_p^L(s)\right|^2 d\widetilde{\mu}_m\Bigg\}^{1/2}\Bigg] \\
&\quad \le c\Bigg[\left|\theta_p^U - \theta_p^L\right| + \sum_{m=1}^K \Bigg\{\int \left|\Lambda_{mp}^U(s) - \Lambda_{mp}^L(s)\right|^2 d\widetilde{\mu}_m\Bigg\}^{1/2} \\
&\quad + \sum_{m=1}^K \Bigg\{\int_0^\tau \left|H_p^U(s) - H_p^L(s)\right|^2 d\widetilde{\mu}_m\Bigg\}^{1/2}\Bigg],
\end{aligned}$$

where c is a constant depending on K, the $L_2(P)$-distance within each bracket pair is $O(\epsilon)$. Hence, the bracket entropy integral of $\mathcal{F}_{3k}$ is finite, so that $\mathcal{F}_{3k}$ is P-Donsker.

Lemma 2. *For any bounded random variable* (θ, Λ) *in* $\Theta \times BV[0,\tau]$*, the function* $g(s) \equiv |E[\dot\Psi_k(\mathcal{O}_i;\theta,\mathcal{A})[I(\cdot \geq s)]/\Psi(\mathcal{O}_i;\theta,\mathcal{A})]|$ *is left-continuous and satisfies that, for any* $s \in [0,\tau]$*, there exist* $\delta_s, c_s > 0$ *such that* $|g(\tilde s) - g(s)| \leq c_s|\tilde s - s|$ *for* $\tilde s \in (s - \delta_s, s)$ *and* $|g(\tilde s) - g(s+)| \leq c_s|\tilde s - s|$ *for* $\tilde s \in (s, s+\delta_s)$*.*

Proof. Since $\mu_{ik}(t;\mathcal{O}_i)$ is non-decreasing in t, it follows from (C4) that for any s_1 and s_2,

$$
\begin{aligned}
|g(s_1) - g(s_2)| &\leq E\left[\mathcal{F}(\mathcal{O}_i)\left\{\int \Big|I(t \geq s_1) - I(t \geq s_2)\Big| d\mu_{ik}(t;\mathcal{O}_i)\right\}\right] \\
&\leq \left|E\Big[\mathcal{F}(\mathcal{O}_i)\mu_{ik}(s_1;\mathcal{O}_i)\Big] - E\Big[\mathcal{F}(\mathcal{O}_i)\mu_{ik}(s_2;\mathcal{O}_i)\Big]\right|.
\end{aligned}
$$

Thus, $g(s)$ is in $BV[0,\tau]$ and is left-continuous. In addition, the left- and right-differentiability of $E[\mathcal{F}(\mathcal{O}_i)\mu_{ik}(s;\mathcal{O}_i)]$ in (C4) implies that the second part of the lemma holds.

Lemma 3. *For any* $h(s) \in BV[0,\tau]$*, the linear map* $h \mapsto \int_0^\tau h(t)\eta_{0km}(t,s;\theta_0,\mathcal{A}_0)d\Lambda_{0k}(t)$ *is a bounded compact operator from* $BV[0,\tau]$ *to* $BV[0,\tau]$*.*

Proof. It is clear from (C6) that this function maps any bounded set in $BV[0,\tau]$ into a bounded set consisting of Lipschitz-continuous functions. The result thus follows since any bounded and Lipschitz-continuous functions consist of a totally bounded set in $BV[0,\tau]$ and the linear map is continuous.

6. Consistency

The following theorem states the consistency of $\widehat\theta$ and $\widehat\Lambda_k$, $k = 1,\ldots,K$.

Theorem 1. *Under* (C1)$-$(C5)*,* $|\widehat\theta - \theta_0| + \sum_{k=1}^K \sup_{t\in[0,\tau]} |\widehat\Lambda_k(t) - \Lambda_{0k}(t)| \to_{a.s.} 0$.

Proof. We fix a random sample in the probability space and assume that (C1)$-$(C5) hold for this sample. The set of such samples has probability one. We prove the result for this fixed sample. The entire proof consists of three steps.

Step 1. We show that the NPMLEs exist or, equivalently, $\widehat\Lambda_k(\tau) < \infty$ $(k = 1,\ldots,K)$ for large n. By (C3), the likelihood function is bounded by

$$
\begin{aligned}
&\prod_{i=1}^n r_1(\mathcal{O}_i)\prod_{k=1}^K\prod_{t\leq\tau}\left[\Lambda_k\{t\}R_{ik\cdot}(t)\left\{1 + \int_0^t R_{ik\cdot}(s)d\Lambda_k(s)\right\}^{-1}\right]^{dN^*_{ik\cdot}(t)} \\
&\qquad \times\left\{1 + \int_0^\tau R_{ik\cdot}(s)d\Lambda_k(s)\right\}^{-c_1}
\end{aligned}
$$

$$\leq \prod_{i=1}^{n} r_1(\mathcal{O}_i) \prod_{k=1}^{K} \left\{1 + \int_0^\tau R_{ik\cdot}(s) d\Lambda_k(s)\right\}^{-c_1}.$$

If $\Lambda_k(\tau) = \infty$ for some k, then (C2) implies that, with probability one, $\inf_{t\in[0,\tau]} R_{ik\cdot}(t) \geq 1$ for some i, so that the above function is equal to zero. Thus, the maximum of the likelihood function can only be attained for $\widehat{\Lambda}_k(\tau) < \infty$.

Step 2. We show that $\limsup_n \widehat{\Lambda}_k(\tau) < \infty$ almost surely, i.e., $\widehat{\Lambda}_k(\tau)$ is bounded uniformly for all large n. By differentiating the objective function (3.1) with respect to $\Lambda_k\{Y_{ikl}\}$ for which $dN^*_{ikl}(Y_{ikl}) = 1$ and $R_{ikl}(Y_{ikl}) = 1$, we note that $\widehat{\Lambda}_k\{Y_{ikl}\}$ satisfies

$$\frac{1}{\widehat{\Lambda}_k\{Y_{ikl}\}} = -\sum_{j=1}^{n} \frac{\dot{\Psi}_k(\mathcal{O}_j; \widehat{\theta}, \widehat{\mathcal{A}})[I(\cdot \geq Y_{ikl})]}{\Psi(\mathcal{O}_j; \widehat{\theta}, \widehat{\mathcal{A}})}.$$

In other words,

$$\widehat{\Lambda}_k(t) = -\sum_{i=1}^{n} \sum_{m=1}^{n_{ik}} \int_0^t \left\{ \sum_{j=1}^{n} \frac{\dot{\Psi}_k(\mathcal{O}_j; \widehat{\theta}, \widehat{\mathcal{A}})[I(\cdot \geq s)]}{\Psi(\mathcal{O}_j; \widehat{\theta}, \widehat{\mathcal{A}})} \right\}^{-1} R_{ikm}(s) dN^*_{ikm}(s).$$

To prove the boundedness of $\widehat{\Lambda}_k(\tau)$, we construct another step function $\widetilde{\Lambda}_k$ with jumps only at the Y_{ikl} for which $dN^*_{ikl}(Y_{ikl}) = 1$ and $R_{ikl}(Y_{ikl}) = 1$,

$$\frac{1}{\widetilde{\Lambda}_k\{Y_{ikl}\}} = -\sum_{j=1}^{n} \frac{\dot{\Psi}_k(\mathcal{O}_j; \theta_0, \mathcal{A}_0)[I(\cdot \geq Y_{ikl})]}{\Psi(\mathcal{O}_j; \theta_0, \mathcal{A}_0)},$$

that is,

$$\widetilde{\Lambda}_k(t) = -\sum_{i=1}^{n} \sum_{m=1}^{n_{ik}} \int_0^t \left\{ \sum_{j=1}^{n} \frac{\dot{\Psi}_k(\mathcal{O}_j; \theta_0, \mathcal{A}_0)[I(\cdot \geq s)]}{\Psi(\mathcal{O}_j; \theta_0, \mathcal{A}_0)} \right\}^{-1} R_{ikm}(s) dN^*_{ikm}(s).$$

We show that $\widetilde{\Lambda}_k$ uniformly converges to Λ_{0k}. By Lemma 1,

$$n^{-1}\left\{ \sum_{j=1}^{n} \frac{\dot{\Psi}_k(\mathcal{O}_j; \theta_0, \mathcal{A}_0)[I(\cdot \geq s)]}{\Psi(\mathcal{O}_j; \theta_0, \mathcal{A}_0)} \right\} \to E\left[\frac{\dot{\Psi}_k(\mathcal{O}_i; \theta_0, \mathcal{A}_0)[I(\cdot \geq s)]}{\Psi(\mathcal{O}_i; \theta_0, \mathcal{A}_0)} \right] \tag{6.1}$$

uniformly in $s \in [0, \tau]$. Since the score function, along the path $\Lambda_k = \Lambda_{0k} + \epsilon I(\cdot \geq s)$ with the other parameters fixed at their true values, has zero expectation,

$$0 = E\left[\sum_{l=1}^{n_{ik}} \int \frac{\delta(t = s)}{\lambda_{0k}(t)} R_{ikl}(t) dN^*_{ikl}(t) \right] + E\left[\frac{\dot{\Psi}_k(\mathcal{O}_i; \theta_0, \mathcal{A}_0)[I(\cdot \geq s)]}{\Psi(\mathcal{O}_i; \theta_0, \mathcal{A}_0)} \right]$$

$$= \frac{E[\sum_{l=1}^{n_{ik}} R_{ikl}(s)dN^*_{ikl}(s)/ds]}{\lambda_{0k}(s)} + E\left[\frac{\dot{\Psi}_k(\mathcal{O}_i;\theta_0,\mathcal{A}_0)[I(\cdot \geq s)]}{\Psi(\mathcal{O}_i;\theta_0,\mathcal{A}_0)}\right], \tag{6.2}$$

where $\delta(t = s)$ is the Dirac function. The submodel is not in the parameter space; however, we can always choose a sequence of submodels in the parameter space which approximates this submodel. Thus, the uniform limit of $\widetilde{\Lambda}_k(t)$ is

$$E\left[\sum_{m=1}^{n_{ik}} \int_0^t \left\{\frac{E[\sum_{l=1}^{n_{ik}} R_{ikl}(s)dN^*_{ikl}(s)/ds]}{\lambda_{0k}(s)}\right\}^{-1} R_{ikm}(s)dN^*_{ikm}(s)\right] = \Lambda_{0k}(t).$$

That is, $\widetilde{\Lambda}_k(t)$ uniformly converges to $\Lambda_{0k}(t)$.

We next show that the difference between the log-likelihood functions evaluated at $(\widehat{\theta},\widehat{\mathcal{A}})$ and $(\theta_0,\widetilde{\mathcal{A}})$, where $\widetilde{\mathcal{A}} = (\widetilde{\Lambda}_1,\ldots,\widetilde{\Lambda}_K)$, is negative eventually if some $\widehat{\Lambda}_k(\tau)$ diverges, which will induce a contradiction. The key arguments are based on (C3). Clearly, $n^{-1}\mathcal{L}_n(\widehat{\theta},\widehat{\mathcal{A}}) \geq n^{-1}\mathcal{L}_n(\theta_0,\widetilde{\mathcal{A}})$. It follows from (6.1) and (6.2) that $n\widetilde{\Lambda}_k\{t\}$ converges to $\lambda_{0k}(t)/E[\sum_{l=1}^{n_{ik}} R_{ikl}(t)dN^*_{ikl}(t)/dt]$, and is thus uniformly bounded away from zero, where t is an observed failure time. Therefore,

$$\begin{aligned}
&n^{-1}\mathcal{L}_n(\theta_0,\widetilde{\mathcal{A}}) + n^{-1}\sum_{i=1}^n\sum_{k=1}^K\sum_{l=1}^{n_{ik}}\int R_{ikl}(t)dN^*_{ikl}(t)\log n\\
&\quad = n^{-1}\sum_{i=1}^n\sum_{k=1}^K\sum_{l=1}^{n_{ik}}\int \log(n\widetilde{\Lambda}_k\{t\})R_{ikl}(t)dN^*_{ikl}(t) + n^{-1}\sum_{i=1}^n \log\Psi(\mathcal{O}_i;\theta_0,\mathcal{A}_0),
\end{aligned}$$

which is bounded away from $-\infty$ when n is large. That is,

$$n^{-1}\mathcal{L}_n(\theta_0,\widetilde{\mathcal{A}}) + n^{-1}\sum_{i=1}^n\sum_{k=1}^K\sum_{l=1}^{n_{ik}}\int R_{ikl}(t)dN^*_{ikl}(t)\log n = O(1),$$

where $O(1)$ denotes a finite constant. On the other hand, (C3) implies that

$$\begin{aligned}
n^{-1}\mathcal{L}_n(\widehat{\theta},\widehat{\mathcal{A}}) &\leq n^{-1}\sum_{i=1}^n\sum_{k=1}^K\sum_{l=1}^{n_{ik}}\int R_{ikl}(t)\log\widehat{\Lambda}_k\{t\}dN^*_{ikl}(t)\\
&\quad + n^{-1}\sum_{i=1}^n \log\Psi(\mathcal{O}_i;\widehat{\theta},\widehat{\mathcal{A}})\\
&\leq n^{-1}\sum_{i=1}^n \log r_1(\mathcal{O}_i) + n^{-1}\sum_{i=1}^n\sum_{k=1}^K\int I(R_{ik\cdot}(t) > 0)\\
&\quad \times \log\widehat{\Lambda}_k\{t\}dN_{ik\cdot}(t)
\end{aligned}$$

$$-n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\int\log\left\{1+\int_0^t R_{ik\cdot}(s)d\widehat{\Lambda}_k(s)\right\}dN_{ik\cdot}(t)$$
$$-n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}c_1\log\left\{1+\int_0^\tau R_{ik\cdot}(s)d\widehat{\Lambda}_k(s)\right\},$$

where $dN_{ik\cdot}(t)=\sum_{l=1}^{n_{ik}}R_{ikl}(t)dN_{ikl}^*(t)$. Thus,

$$\begin{aligned}O(1)\leq n^{-1}&\sum_{i=1}^{n}\sum_{k=1}^{K}\int I(R_{ik\cdot}(t)>0)\log(n\widehat{\Lambda}_k\{t\})dN_{ik\cdot}(t)\\&-n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\int\log\left\{1+\int_0^t R_{ik\cdot}(s)d\widehat{\Lambda}_k(s)\right\}dN_{ik\cdot}(t)\\&-n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}c_1\log\left\{1+\int_0^\tau R_{ik\cdot}(s)d\widehat{\Lambda}_k(s)\right\}.\end{aligned}\tag{6.3}$$

We now show that the right-hand side diverges to $-\infty$ if $\widehat{\Lambda}_k(\tau)$ diverges for some k. The proof is based on the partitioning idea of Murphy (1994). Specifically, we construct a sequence $t_{0k}=\tau>t_{1k}>t_{2k}>\ldots$ in the following manner. First, we define

$$\begin{aligned}t_{1k}=\operatorname{argmin}\Bigg\{&t\in[0,t_{0k}):\frac{c_1}{2}E[I(\overline{R}_{ik\cdot}(\tau)>0)]\\&\geq E\left[I(\overline{R}_{ik\cdot}(t)>0,\overline{R}_{ik\cdot}(\tau)=0)\int_t^{t_{0k}}dN_{ik\cdot}(t)\right]\Bigg\},\end{aligned}$$

where $\overline{R}_{ik\cdot}(t)=\inf_{s\in[0,t]}R_{ik\cdot}(s)$. Clearly, such a t_{1k} exists, and the above inequality becomes an equality if $t_{1k}>0$. If $t_{1k}>0$, we choose a small constant ϵ_0 such that

$$\frac{\epsilon_0}{1-\epsilon_0}<\frac{c_1E[I(\overline{R}_{ik\cdot}(\tau)=0,\overline{R}_{ik\cdot}(t_{1k})>0)}{E[I(\overline{R}_{ik\cdot}(t_{1k})=0,\overline{R}_{ik\cdot}(0)>0)\int_0^\tau dN_{ik\cdot}(t)]},$$

and define

$$\begin{aligned}t_{2k}=\operatorname{argmin}\Bigg\{&t\in[0,t_{1k}):(1-\epsilon_0)E\Bigg[\left\{c_1+\int_{t_{1k}}^{t_{0k}}dN_{ik\cdot}(t)\right\}\\&\times I\Big(\overline{R}_{ik\cdot}(t_{0k})=0,\overline{R}_{ik\cdot}(t_{1k})>0\Big)\Bigg]\\&\geq E\left[I\Big(\overline{R}_{ik\cdot}(t_{1k})=0,\overline{R}_{ik\cdot}(t)>0\Big)\int_t^{t_{1k}}dN_{ik\cdot}(t)\right]\Bigg\}.\end{aligned}$$

Such a t_{2k} exists. If $t_{2k} > 0$, the inequality is an equality, and we define

$$\begin{aligned} t_{3k} = \operatorname{argmin}\Bigg\{ t \in [0, t_{1k}) : (1-\epsilon_0) E\Bigg[\Big\{ c_1 + \int_{t_{2k}}^{t_{1k}} dN_{ik\cdot}(t) \Big\} \\ \times I\Big(\overline{R}_{ik\cdot}(t_{1k}) = 0, \overline{R}_{ik\cdot}(t_{2k}) > 0\Big)\Bigg] \\ \geq E\Big[I\Big(\overline{R}_{ik\cdot}(t_{2k}) = 0, \overline{R}_{ik\cdot}(t) > 0\Big) \int_t^{t_{2k}} dN_{ik\cdot}(t)\Big]\Bigg\}. \end{aligned}$$

We continue this process. The sequence eventually stops at some $t_{N_k,k} = 0$. If this is not true, then the sequence is infinite and strictly decreases to some $t^* \geq 0$. Since all the inequalities are equalities, we sum all the equations except the first one to obtain

$$\begin{aligned} (1-\epsilon_0) E\Bigg[\Big\{ c_1 + \int_{t^*}^{t_{0k}} dN_{ik\cdot}(t) \Big\} I\Big(\overline{R}_{ik\cdot}(t^*) > 0, \overline{R}_{ik\cdot}(\tau) = 0\Big)\Bigg] \\ = E\Big[I\Big(\overline{R}_{ik\cdot}(t_{1k}) = 0, \overline{R}_{ik\cdot}(t^*) > 0\Big) \int_{t^*}^{t_{1k}} dN_{ik\cdot}(t)\Big], \end{aligned}$$

which implies that

$$\begin{aligned} c_1(1-\epsilon_0) E\Big[I(\overline{R}_{ik\cdot}(\tau) = 0, \overline{R}_{ik\cdot}(t_{1k}) > 0)\Big] \\ \leq \epsilon_0 E\Big[I\Big(\overline{R}_{ik\cdot}(t_{1k}) = 0, \overline{R}_{ik\cdot}(0) > 0\Big) \int_0^{\tau} dN_{ik\cdot}(t)\Big]. \end{aligned}$$

This contradicts the choice of ϵ_0. Thus, the sequence stops at some $t_{N_k k} = 0$.

If we write $I_{qk} = [t_{q+1,k}, t_{qk})$, then the right-hand side of (6.3) can be bounded by

$$\begin{aligned} \sum_{k=1}^{K} \Bigg[n^{-1} \sum_{i=1}^{n} \sum_{q=0}^{N_k-1} I\Big(\overline{R}_{ik\cdot}(t_{qk}) = 0, \overline{R}_{ik\cdot}(t_{q+1,k}) > 0\Big) \\ \times \int_{t \in I_{qk}} \log\Big(n\widehat{\Lambda}_k\{t\}\Big) dN_{ik\cdot} \\ - n^{-1} \sum_{i=1}^{n} \sum_{q=0}^{N_k-1} I\Big(\overline{R}_{ik\cdot}(t_{qk}) = 0, \overline{R}_{ik\cdot}(t_{q+1,k}) > 0\Big) \\ \times \int_{t \in I_{qk}} dN_{ik\cdot} \log\Big\{1 + \widehat{\Lambda}_k(t_{q+1,k})\Big\} \\ - n^{-1} \sum_{i=1}^{n} \sum_{q=0}^{N_k-1} I\Big(\overline{R}_{ik\cdot}(t_{qk}) = 0, \overline{R}_{ik\cdot}(t_{q+1,k}) > 0\Big) c_1 \log\Big\{1 + \widehat{\Lambda}_k(t_{q+1,k})\Big\} \end{aligned}$$

$$-n^{-1}\sum_{i=1}^{n} I\Big(\overline{R}_{ik\cdot}(t_{0k})>0\Big)\log\Big\{1+\widehat{\Lambda}_k(\tau)\Big\}\Bigg]. \qquad (6.4)$$

Since $\log x$ is a concave function,

$$\begin{aligned}
&\sum_{i=1}^{n} I\Big(\overline{R}_{ik\cdot}(t_{qk})=0,\overline{R}_{ik\cdot}(t_{q+1,k})>0\Big)\int_{t\in I_{qk}}\log\Big(n\widehat{\Lambda}_k\{t\}\Big)dN_{ik\cdot}(t)\\
&\quad\le\Big\{\sum_{i=1}^{n} I\Big(\overline{R}_{ik\cdot}(t_{qk})=0,\overline{R}_{ik\cdot}(t_{q+1,k})>0\Big)\int_{t\in I_{qk}}dN_{ik\cdot}\Big\}\\
&\qquad\times\log\left[\frac{\sum_{i=1}^{n} I(\overline{R}_{ik\cdot}(t_{qk})=0,\overline{R}_{ik\cdot}(t_{q+1,k})>0)\int_{t\in I_{qk}}n\widehat{\Lambda}_k\{t\}dN_{ik\cdot}(t)}{\sum_{i=1}^{n} I(\overline{R}_{ik\cdot}(t_{qk})=0,\overline{R}_{ik\cdot}(t_{q+1,k})>0)\int_{t\in I_{qk}}dN_{ik\cdot}(t)}\right]\\
&\quad\le\Big\{\sum_{i=1}^{n} I\Big(\overline{R}_{ik\cdot}(t_{qk})=0,\overline{R}_{ik\cdot}(t_{q+1,k})>0\Big)\int_{t\in I_{qk}}dN_{ik\cdot}\Big\}\\
&\qquad\times\log\left[\frac{n\widehat{\Lambda}_k(t_{qk})}{\sum_{i=1}^{n} I(\overline{R}_{ik\cdot}(t_{qk})=0,\overline{R}_{ik\cdot}(t_{q+1,k})>0)\int_{t\in I_{qk}}dN_{ik\cdot}(t)}\right].
\end{aligned}$$

Therefore, (6.4) can be further bounded by

$$\begin{aligned}
O(1)\le&\sum_{k=1}^{K}\Bigg[\sum_{q=0}^{N_k-1}n^{-1}\sum_{i=1}^{n} I\Big(\overline{R}_{ik\cdot}(t_{qk})=0,\overline{R}_{ik\cdot}(t_{q+1,k})>0\Big)\int_{t\in I_{qk}}dN_{ik\cdot}\\
&\times\log\left\{\frac{n}{\sum_{i=1}^{n} I(\overline{R}_{ik\cdot}(t_{qk})=0,\overline{R}_{ik\cdot}(t_{q+1,k})>0)\int_{t\in I_{qk}}dN_{ik\cdot}}\right\}\\
&+\sum_{q=0}^{N_k-1}\log\widehat{\Lambda}_k(t_{qk})\Big\{n^{-1}\sum_{i=1}^{n} I\Big(\overline{R}_{ik\cdot}(t_{qk})=0,\overline{R}_{ik\cdot}(t_{q+1,k})>0\Big)\int_{t\in I_{qk}}dN_{ik\cdot}\Big\}\\
&-n^{-1}\sum_{i=1}^{n}\sum_{q=0}^{N_k-1} I\Big(\overline{R}_{ik\cdot}(t_{qk})=0,\overline{R}_{ik\cdot}(t_{q+1,k})>0\Big)\\
&\times\int_{t\in I_{qk}}dN_{ik\cdot}\log\Big\{1+\widehat{\Lambda}_k(t_{q+1,k})\Big\}\Big\}\\
&-\sum_{q=0}^{N_k-1}n^{-1}\sum_{i=1}^{n} I\Big(\overline{R}_{ik\cdot}(t_{qk})=0,\overline{R}_{ik\cdot}(t_{q+1,k})>0\Big)c_1\log\Big\{1+\widehat{\Lambda}_k(t_{q+1,k})\Big\}\Big\}\\
&-n^{-1}\sum_{i=1}^{n} I\Big(\overline{R}_{ik\cdot}(t_{0k})>0\Big)\log\Big\{1+\widehat{\Lambda}_k(\tau)\Big\}\Bigg].
\end{aligned}$$

By (C2),

$$\frac{n}{\sum_{i=1}^{n} I(\overline{R}_{ik\cdot}(t_{qk})=0,\overline{R}_{ik\cdot}(t_{q+1,k})>0)\int_{t\in I_{qk}} dN_{ik\cdot}}$$
$$\to_{a.s.} \left(E\Big[I\Big(\overline{R}_{ik\cdot}(t_{qk})=0,\overline{R}_{ik\cdot}(t_{q+1,k})>0\Big)\int_{t\in I_{qk}} dN_{ik\cdot}\Big]\right)^{-1}<\infty,$$

so that

$$\begin{aligned}
O(1)\le \sum_{k=1}^{K}\Bigg(&-n^{-1}\sum_{i=1}^{n}\frac{c_1}{2}I\Big(\overline{R}_{ik\cdot}(t_{0k})>0\Big)\log\Big\{1+\widehat{\Lambda}_k(\tau)\Big\}\\
&-\Big\{n^{-1}\sum_{i=1}^{n}\frac{c_1}{2}I\Big(\overline{R}_{ik\cdot}(t_{0k})>0\Big)-n^{-1}\sum_{i=1}^{n}I\Big(\overline{R}_{ik\cdot}(t_{0k})=0,\overline{R}_{ik\cdot}(t_{1k})>0\Big)\\
&\times\int_{t\in I_{0k}} dN_{ik\cdot}\Big\}\log\Big\{1+\widehat{\Lambda}_k(t_{0k})\Big\}\\
&-\sum_{q=1}^{N_k-1}\Big[n^{-1}\sum_{i=1}^{n}I\big(\overline{R}_{ik\cdot}(t_{q-1,k})=0,\overline{R}_{ik\cdot}(t_{qk})>0\big)\Big\{c_1+\int_{t\in I_{qk}} dN_{ik\cdot}\Big\}\\
&-n^{-1}\sum_{i=1}^{n}I\Big(\overline{R}_{ik\cdot}(t_{qk})=0,\overline{R}_{ik\cdot}(t_{q+1,k})>0\Big)\int_{t\in I_{qk}} dN_{ik\cdot}\Big]\\
&\times\Big\{1+\log\widehat{\Lambda}_k(t_{qk})\Big\}\Bigg).
\end{aligned}$$

According to the construction of the t_{qk}'s, the coefficients in front of $\log\widehat{\Lambda}_k(t_{qk})$ are all negative when n is large enough. Therefore, the corresponding terms cannot diverge to ∞. However, if $\widehat{\Lambda}_k(\tau)\to\infty$, the first term in the summation goes to $-\infty$. We conclude that for all n large enough, $\widehat{\Lambda}_k(\tau)<\infty$. Thus, $\limsup_n \widehat{\Lambda}_k(\tau)<\infty$.

Step 3. We obtain the consistency result from (C5). Since $\widehat{\Lambda}_k$ is bounded and monotone, $\widehat{\Lambda}_k$ is weakly compact. Helly's Selection Theorem implies that, for any subsequence, we can always choose a further subsequence such that $\widehat{\Lambda}_k$ point-wise converges to some monotone function Λ_k^*. Without loss of generality, we also assume that $\widehat{\theta}$ converges to some θ^*. The consistency will hold if we can show that $\Lambda_k^*=\Lambda_{0k}$ and $\theta^*=\theta_0$. Since Λ_{0k} is continuous, the weak convergence of $\widehat{\Lambda}_k$ to Λ_{0k} can be strengthened to the uniform convergence of $\widehat{\Lambda}_k$ to Λ_{0k} in $[0,\tau]$.

Note that

$$\widehat{\Lambda}_k(t)=\int_0^t \frac{|n^{-1}\sum_{j=1}^{n}\dot{\Psi}_k(\mathcal{O}_j;\theta_0,\mathcal{A}_0)[I(\cdot\ge s)]/\Psi(\mathcal{O}_j;\theta_0,\mathcal{A}_0)|}{|n^{-1}\sum_{j=1}^{n}\dot{\Psi}_k(\mathcal{O}_j;\widehat{\theta},\widehat{\mathcal{A}})[I(\cdot\ge s)]/\Psi(\mathcal{O}_j;\widehat{\theta},\widehat{\mathcal{A}})|}d\widetilde{\Lambda}_k(s). \tag{6.5}$$

Clearly, $\widehat{\Lambda}_k$ is absolutely continuous with respect to $\widetilde{\Lambda}_k$. By condition (C3),

$$\sup_{s\in[0,\tau]}\left|n^{-1}\sum_{j=1}^{n}\frac{\dot{\Psi}_k(\mathcal{O}_j;\widehat{\theta},\widehat{\mathcal{A}})[I(\cdot\geq s)]}{\Psi(\mathcal{O}_j;\widehat{\theta},\widehat{\mathcal{A}})}-n^{-1}\sum_{j=1}^{n}\frac{\dot{\Psi}_k(\mathcal{O}_j;\theta^*,\mathcal{A}^*)[I(\cdot\geq s)]}{\Psi(\mathcal{O}_j;\theta^*,\mathcal{A}^*)}\right|$$
$$\leq n^{-1}\sum_{j=1}^{n}\mathcal{F}(\mathcal{O}_j)\left\{|\widehat{\theta}-\theta^*|+\sum_{k=1}^{K}\int|\widehat{\Lambda}_k(t)-\Lambda_k^*(t)|d\mu_{jk}(t;\mathcal{O}_j)\right\}\to 0,$$

since $\widehat{\Lambda}_k$ converges to Λ_k^* and is bounded and $\{\mathcal{F}(\mathcal{O}_j)\mu_{jk}(t;\mathcal{O}_j):t\in[0,\tau]\}$ is a P-Glivenko-Cantelli class. By Lemma 1 and the Glivenko-Cantelli Theorem,

$$n^{-1}\sum_{j=1}^{n}\frac{\dot{\Psi}_k(\mathcal{O}_j;\theta^*,\mathcal{A}^*)[I(\cdot\geq s)]}{\Psi(\mathcal{O}_j;\theta^*,\mathcal{A}^*)}$$
$$\to E\left[\frac{\dot{\Psi}_k(\mathcal{O}_j;\theta^*,\mathcal{A}^*)[I(\cdot\geq s)]}{\Psi(\mathcal{O}_j;\theta^*,\mathcal{A}^*)}\right]\quad\text{uniformly in } s\in[0,\tau],$$

$$n^{-1}\sum_{j=1}^{n}\frac{\dot{\Psi}_k(\mathcal{O}_j;\theta_0,\mathcal{A}_0)[I(\cdot\geq s)]}{\Psi(\mathcal{O}_j;\theta_0,\mathcal{A}_0)}$$
$$\to E\left[\frac{\dot{\Psi}_k(\mathcal{O}_j;\theta_0,\mathcal{A}_0)[I(\cdot\geq s)]}{\Psi(\mathcal{O}_j;\theta_0,\mathcal{A}_0)}\right]\quad\text{uniformly in } s\in[0,\tau].$$

The numerator and denominator in the integrand of (6.5) converge uniformly to deterministic functions, denoted by $g_{1k}(s)$ and $g_{2k}(s)$, respectively. It follows from (6.2) that $g_{1k}(s)\equiv E[\sum_{l=1}^{n_{ik}}R_{ikl}(s)dN_{ikl^*(s)}/ds]/\lambda_{ik}(s)$ is bounded away from zero. We claim that $\inf_{s\in[0,\tau]}g_{2k}(s)>0$. If this is not true, then there exists some $s^*\in[0,\tau]$ such that $g_{2k}(s^*+)=0$ or $g_{2k}(s^*)=0$. By Lemma 2, there exist δ^* and c^* such that $|g_{2k}(s)|\leq c^*|s-s^*|$ for $s\in(s^*,s^*+\delta^*)$ or $s\in(s^*-\delta^*,s^*]$. On the other hand, for any $\epsilon>0$,

$$\widehat{\Lambda}_k(\tau)\geq\int_0^\tau\frac{|n^{-1}\sum_{j=1}^{n}\dot{\Psi}_k(\mathcal{O}_j;\theta_0,\mathcal{A}_0)[I(\cdot\geq s)]/\Psi(\mathcal{O}_j;\theta_0,\mathcal{A}_0)|}{\epsilon+|n^{-1}\sum_{j=1}^{n}\dot{\Psi}_k(\mathcal{O}_j;\widehat{\theta},\widehat{\mathcal{A}})[I(\cdot\geq s)]/\Psi(\mathcal{O}_j;\widehat{\theta},\widehat{\mathcal{A}})|}d\widetilde{\Lambda}_k(s).$$

Taking limits on both sides, we obtain $O(1)\geq\int_0^\tau\{\epsilon+g_{2k}(s)\}^{-1}g_{1k}(s)d\Lambda_{0k}(s)$. Let $\epsilon\to 0$. By the Monotone Convergence Theorem, $O(1)\geq\int_{s^*}^{s^*+\delta^*}\{c^*|s-s^*|\}^{-1}g_{1k}(s)\lambda_{0k}(s)ds$, or $O(1)\geq\int_{s^*-\delta^*}^{s^*}\{c^*|s-s^*|\}^{-1}g_{1k}(s)\lambda_{0k}(s)ds$. This is a contradiction since the right-hand side is infinite. The contradiction implies that the limit $g_{2k}(s)$ is uniformly positive. We can take limits on both sides of (6.5) to obtain $\Lambda_k^*(t)=\int_0^t g_{2k}^{-1}(s)g_{1k}(s)d\Lambda_{0k}(s)$. Thus, Λ_k^* is also absolutely continuous with respect to Λ_{0k} and $d\Lambda_k^*/d\Lambda_{0k}=g_{1k}/g_{2k}$. Since $\Lambda_{0k}(t)$ is differentiable with respect to t, so is $\Lambda_k^*(t)$. We write $\{\Lambda_k^*\}'(t) = \lambda_k^*(t)$. The forgoing arguments show that $d\widehat{\Lambda}_k(t)/d\widetilde{\Lambda}_k(t)$ uniformly converges to $\lambda_k^*(t)/\lambda_{0k}(t)$, which is uniformly positive in $[0, \tau]$.

It follows from the inequality $n^{-1}\mathcal{L}_n(\widehat{\theta}, \widehat{\mathcal{A}}) \geq n^{-1}\mathcal{L}_n(\theta_0, \widetilde{\mathcal{A}})$ that

$$n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{l=1}^{n_{ik}}\int \log\frac{d\widehat{\Lambda}_k(t)}{d\widetilde{\Lambda}_k(t)} R_{ikl}(t)dN_{ikl}^*(t) + n^{-1}\sum_{i=1}^{n}\log\frac{\Psi(\mathcal{O}_i;\widehat{\theta},\widehat{\mathcal{A}})}{\Psi(\mathcal{O}_i;\theta_0,\widetilde{\mathcal{A}})} \geq 0.$$

In view of Lemma 1, the Glivenko-Cantelli Theorem and the uniform convergence of $d\widehat{\Lambda}_k/d\widetilde{\Lambda}_k$, taking limits on both sides of the above inequality yields

$$E\left[\log\frac{\prod_{k=1}^{K}\prod_{l=1}^{n_{ik}}\prod_{t\leq\tau}\{\lambda_k^*(t)\}^{R_{ikl}(t)dN_{ikl}^*(t)}\Psi(\mathcal{O}_i;\theta^*,\mathcal{A}^*)}{\prod_{k=1}^{K}\prod_{l=1}^{n_{ik}}\prod_{t\leq\tau}\{\lambda_0(t)\}^{R_{ikl}(t)dN_{ikl}^*(t)}\Psi(\mathcal{O}_i;\theta_0,\mathcal{A}_0)}\right] \geq 0.$$

The left-hand side is the negative Kullback-Leibler distance of the density indexed by $(\theta^*, \mathcal{A}^*)$. Thus, (C5) entails that $\theta^* = \theta_0$ and $\Lambda^* = \Lambda_0$.

7. Weak Convergence and Asymptotic Efficiency

Define $\mathcal{V} = \{v \in R^d, |v| \leq 1\}$, and $\mathcal{Q} = \{h(t) : \|h(t)\|_{V[0,\tau]} \leq 1\}$. We identify $(\widehat{\theta} - \theta_0, \widehat{\mathcal{A}} - \mathcal{A}_0)$ as a random element in $l^\infty(\mathcal{V} \times \mathcal{Q}^K)$ through the definition $(\widehat{\theta} - \theta_0)^T v + \sum_{k=1}^K \int_0^\tau h_k(s)d(\widehat{\Lambda}_k - \Lambda_{0k})(s)$.

Theorem 2. *Under* (C1)–(C7), $n^{1/2}(\widehat{\theta} - \theta_0, \widehat{\mathcal{A}} - \mathcal{A}_0) \to_d \mathcal{G}$ *in* $l^\infty(\mathcal{V} \times \mathcal{Q}^K)$, *where* $\mathcal{G}$ *is a continuous zero-mean Gaussian process. Furthermore, the limiting covariance matrix of* $n^{1/2}(\widehat{\theta} - \theta_0)$ *attains the semiparametric efficiency bound.*

Proof. The proof is based on the likelihood equation and follows the arguments of Van der Vaart (1998, pp.419-424). Let $\mathcal{L}(\theta, \mathcal{A})$ be the log-likelihood function from a single cluster, $\dot{\mathcal{L}}_\theta(\theta, \mathcal{A})$ be the derivative of $\mathcal{L}(\theta, \mathcal{A})$ with respect to θ, and $\dot{\mathcal{L}}_k(\theta, \mathcal{A})[H_k]$ be the path-wise derivative along the path $\Lambda_k + \epsilon H_k$. We sometimes omit the arguments in these derivatives when $\theta = \theta_0$ and $\mathcal{A} = \mathcal{A}_0$. Let $\mathcal{P}_n$ be the empirical measure based on n i.i.d. observations, and $\mathcal{P}$ be its expectation.

Let $\mathcal{W} = (h_1, \ldots, h_K) \in \mathcal{Q}^K$. The likelihood equation for $(\widehat{\theta}, \widehat{\mathcal{A}})$ along the path $(\widehat{\theta} + \epsilon v, \widehat{\mathcal{A}} + \epsilon\int \mathcal{W}d\widehat{\mathcal{A}})$, where $v \in R^d$ and $h_k \in BV[0, \tau]$, is given by

$$0 = \mathcal{P}_n\left[v^T\dot{\mathcal{L}}_\theta(\theta, \mathcal{A}) + \sum_{k=1}^{K}\dot{\mathcal{L}}_k(\theta, \mathcal{A})\left[\int h_k d\Lambda_k\right]\right].$$

To be specific,

$$0 = \mathcal{P}_n\left[\frac{v^T\dot{\Psi}_\theta(\mathcal{O}_i;\theta,\mathcal{A})}{\Psi(\mathcal{O}_i;\theta,\mathcal{A})}\right] + \sum_{k=1}^{K}\mathcal{P}_n\left[\sum_{l=1}^{n_{ik}}\int h_k(t)R_{ikl}(t)dN_{ikl}^*(t)\right.$$

$$+\dot{\Psi}_k(\mathcal{O}_i;\theta,\mathcal{A})\Big[\int h_k d\Lambda_k\Big]\Big].$$

Since $(\theta_0,\mathcal{A}_0)$ maximizes $\mathcal{P}[\mathcal{L}(\theta,\mathcal{A})]$,

$$0=\mathcal{P}\Big[v^T\dot{\mathcal{L}}_\theta(\theta_0,\mathcal{A}_0)\Big],\quad 0=\mathcal{P}\Big[\dot{\mathcal{L}}_k(\theta_0,\mathcal{A}_0)\Big[\int h_k d\Lambda_{0k}\Big]\Big],\quad h_k\in\mathcal{Q},\quad k=1,\ldots,K.$$

These equations, combined with the likelihood equation for $(\widehat{\theta},\widehat{\mathcal{A}})$, yield

$$\begin{aligned}
&n^{1/2}(\mathcal{P}_n-\mathcal{P})\Big[v^T\dot{\mathcal{L}}_\theta(\widehat{\theta},\widehat{\mathcal{A}})+\sum_{k=1}^K\mathcal{L}_k(\widehat{\theta},\widehat{\mathcal{A}})\Big[\int h_k d\widehat{\Lambda}_k\Big]\Big]\\
&\quad=-n^{1/2}\mathcal{P}\Big[\frac{v^T\dot{\Psi}_\theta(\mathcal{O}_i;\widehat{\theta},\widehat{\mathcal{A}})}{\Psi(\mathcal{O}_i;\widehat{\theta},\widehat{\mathcal{A}})}-\frac{v^T\dot{\Psi}_\theta(\mathcal{O}_i;\theta_0,\mathcal{A}_0)}{\Psi(\mathcal{O}_i;\theta_0,\mathcal{A}_0)}\Big]\\
&\qquad-\sum_{k=1}^K n^{1/2}\mathcal{P}\Big[\frac{\dot{\Psi}_k(\mathcal{O}_i;\widehat{\theta},\widehat{\mathcal{A}})[\int h_k d\widehat{\Lambda}_k]}{\Psi(\mathcal{O}_i;\widehat{\theta},\widehat{\mathcal{A}})}-\frac{\dot{\Psi}_k(\mathcal{O}_i;\theta_0,\mathcal{A}_0)[\int h_k d\Lambda_{0k}]}{\Psi(\mathcal{O}_i;\theta_0,\mathcal{A}_0)}\Big].
\end{aligned}$$

Define $\mathcal{N}_0=\{(\theta,\mathcal{A}):|\theta-\theta_0|+\sum_{k=1}^K\|\Lambda_k-\Lambda_{0k}\|_{V[0,\tau]}<\delta_0\}$, where δ_0 is a small positive constant. When n is large enough, $(\widehat{\theta},\widehat{\mathcal{A}})$ belongs to $\mathcal{N}_0$ with probability one. By Lemma 1 and the Donsker Theorem,

$$\begin{aligned}
&o_p(1)+n^{1/2}(\mathcal{P}_n-\mathcal{P})\Big[v^T\dot{\mathcal{L}}_\theta(\theta_0,\mathcal{A}_0)+\sum_{k=1}^K\mathcal{L}_k(\theta_0,\mathcal{A}_0)\Big[\int h_k d\Lambda_{0k}\Big]\Big]\\
&\quad=-n^{1/2}\mathcal{P}\Big[\frac{v^T\dot{\Psi}_\theta(\mathcal{O}_i;\widehat{\theta},\widehat{\mathcal{A}})}{\Psi(\mathcal{O}_i;\widehat{\theta},\widehat{\mathcal{A}})}-\frac{v^T\dot{\Psi}_\theta(\mathcal{O}_i;\theta_0,\mathcal{A}_0)}{\Psi(\mathcal{O}_i;\theta_0,\mathcal{A}_0)}\Big]\\
&\qquad-\sum_{k=1}^K n^{1/2}\mathcal{P}\Big[\frac{\dot{\Psi}_k(\mathcal{O}_i;\widehat{\theta},\widehat{\mathcal{A}})[\int h_k d\widehat{\Lambda}_k]}{\Psi(\mathcal{O}_i;\widehat{\theta},\widehat{\mathcal{A}})}-\frac{\dot{\Psi}_k(\mathcal{O}_i;\theta_0,\mathcal{A}_0)[\int h_k d\Lambda_{0k}]}{\Psi(\mathcal{O}_i;\theta_0,\mathcal{A}_0)}\Big],
\end{aligned}\tag{7.1}$$

where $o_p(1)$ represents some random element converging in probability to zero in $l^\infty(\mathcal{V}\times\mathcal{Q}^K)$.

Under (C6), the first term on the right-hand side of (7.1) is

$$\begin{aligned}
&-n^{1/2}\Big\{\sum_{k=1}^K\int_0^\tau v^T\zeta_{0k}(s)d(\widehat{\Lambda}_k-\Lambda_{0k})+v^T\zeta_{0\theta}(\widehat{\theta}-\theta_0)\Big\}\\
&\qquad+o\Big(n^{1/2}|\widehat{\theta}-\theta_0|+n^{1/2}\sum_{k=1}^K\|\widehat{\Lambda}_k-\Lambda_{0k}\|_{V[0,\tau]}\Big).
\end{aligned}$$

The second term is $-\sum_{k=1}^K n^{1/2}\{\int_0^\tau h_k(t)\eta_{0k}(t;\widehat{\theta},\widehat{\mathcal{A}})d\widehat{\Lambda}_k(t) - \int_0^\tau h_k(y)\eta_{0k}(t;\theta_0, \mathcal{A}_0)d\Lambda_{0k}(t)\}$. It follows from (C6) that the above expression is

$$
\begin{aligned}
&-\sum_{k=1}^K n^{1/2}\Bigg[\int_0^\tau h_k(t)\Bigg\{\eta_{0k\theta}(t;\theta_0,\mathcal{A}_0)(\widehat{\theta}-\theta_0)\\
&\quad+\sum_{m=1}^K\int_0^\tau \eta_{0km}(s,t;\theta_0,\mathcal{A}_0)d(\widehat{\Lambda}_m-\Lambda_{0m})(s)\Bigg\}d\Lambda_{0k}(t)\\
&\quad+\int_0^\tau h_k(t)\eta_{0k}(t;\theta_0,\mathcal{A}_0)d(\widehat{\Lambda}_k(t)-\Lambda_{0k}(t))\Bigg]\\
&\quad+o\Bigg(n^{1/2}|\widehat{\theta}-\theta_0|+n^{1/2}\sum_{k=1}^K\|\widehat{\Lambda}_k-\Lambda_{0k}\|_{V[0,\tau]}\Bigg)\\
&=-\sum_{k=1}^K n^{1/2}\Bigg[(\widehat{\theta}-\theta_0)^T\int_0^\tau h_k(t)\eta_{0k\theta}(t;\theta_0,\mathcal{A}_0)d\Lambda_{0k}(t)\\
&\quad+\sum_{m=1}^K\int_0^\tau\Bigg\{I(m=k)h_m(t)\eta_{0m}(t;\theta_0,\mathcal{A}_0)\\
&\quad+\int_0^\tau \eta_{0km}(s,t;\theta_0,\mathcal{A}_0)h_k(s)d\Lambda_{0k}(s)\Bigg\}d(\widehat{\Lambda}_m(t)-\Lambda_{0m}(t))\Bigg]\\
&\quad+o\Bigg(n^{1/2}|\widehat{\theta}-\theta_0|+n^{1/2}\sum_{k=1}^K\|\widehat{\Lambda}_k-\Lambda_{0k}\|_{V[0,\tau]}\Bigg).
\end{aligned}
$$

Thus, the right-hand side of (7.1) can be written as

$$
\begin{aligned}
&-n^{1/2}\Bigg\{B_1[v,\mathcal{W}]^T(\widehat{\theta}-\theta_0)+\sum_{k=1}^K\int B_{2k}[v,\mathcal{W}]d(\widehat{\Lambda}_k-\Lambda_{0k})\Bigg\}\\
&\qquad+o\Bigg(n^{1/2}|\widehat{\theta}-\theta_0|+n^{1/2}\sum_{k=1}^K\|\widehat{\Lambda}_k-\Lambda_{0k}\|_{V[0,b\tau]}\Bigg),
\end{aligned}
$$

where $(B_1, B_{21}, \ldots, B_{2K})$ are linear operators in $R^d \times \{BV[0,\tau]\}^K$, and

$$
B_1[v,\mathcal{W}] = v^T\zeta_{0\theta}(\theta_0,\mathcal{A}_0)+\sum_{k=1}^K\int_0^\tau h_k(t)\eta_{0k\theta}(t;\theta_0,\mathcal{A}_0)d\Lambda_{0k}(t), \tag{7.2}
$$

$$
\begin{aligned}
B_{2k}[v,\mathcal{W}] &= v^T\zeta_{0k}(s;\theta_0,\mathcal{A}_0)+h_k(t)\eta_{0k}(t;\theta_0,\mathcal{A}_0)\\
&\quad+\sum_{m=1}^K\int_0^\tau \eta_{0mk}(s,t;\theta_0,\mathcal{A}_0)h_m(s)d\Lambda_{0k}(s),\quad k=1,\ldots,K.
\end{aligned} \tag{7.3}
$$

It follows from the above derivation that

$$
\begin{aligned}
&B_1[v,\mathcal{W}]^T\tilde{v}+\sum_{k=1}^{K}\int B_{2k}[v,\mathcal{W}]\widetilde{W}_k d\Lambda_{0k}\\
&=\frac{d}{d\epsilon}\Big|_{\epsilon=0}\mathcal{P}\Bigg[v^T\mathcal{L}_\theta\Big(\theta_0+\epsilon\tilde{v},\mathcal{A}_0+\epsilon\int\widetilde{\mathcal{W}}d\mathcal{A}_0\Big)\\
&\quad+\sum_{k=1}^{K}\mathcal{L}_k\Big(\theta_0+\epsilon\tilde{v},\mathcal{A}_0+\epsilon\int\widetilde{\mathcal{W}}d\mathcal{A}_0\Big)\Big[\int h_k d\Lambda_{0k}\Big]\Bigg]. \qquad (7.4)
\end{aligned}
$$

We can write $(B_1, B_{21},\ldots,B_{2K})[v,\mathcal{W}]$ as

$$
\begin{pmatrix} v \\ \eta_{01}(t;\theta_0,\mathcal{A}_0)\times h_1(t) \\ \vdots \\ \eta_{0K}(t;\theta_0,\mathcal{A}_0)\times h_K(t)\end{pmatrix}+\begin{pmatrix} v^T\zeta_{0\theta}(\theta_0,\mathcal{A}_0)+\sum_{k=1}^K\int_0^\tau h_k(t)\eta_{0k\theta}(t;\theta_0,\mathcal{A}_0)d\Lambda_{0k}(t)-v \\ v^T\zeta_{01}(t;\theta_0,\mathcal{A}_0)+\sum_{m=1}^K\int_0^\tau \eta_{0m1}(s,t;\theta_0,\mathcal{A}_0)h_m(s)d\Lambda_{0m}(s) \\ \vdots \\ v^T\zeta_{0K}(t;\theta_0,\mathcal{A}_0)+\sum_{m=1}^K\int_0^\tau \eta_{0mK}(s,t;\theta_0,\mathcal{A}_0)h_m(s)d\Lambda_{0m}(s)\end{pmatrix}.
$$

We wish to prove that $(B_1, B_{21},\ldots,B_{2K})$ is invertible. As shown at the end of this section, $\eta_{0k}(t;\theta_0,\mathcal{A}_0)<0$, so that the first term of $(B_1, B_{21},\ldots,B_{2K})$ is an invertible operator. It follows from Lemma 3 that the second term is a compact operator. Thus, $(B_1, B_{21},\ldots,B_{2K})$ is a Fredholm operator, and the invertibility of $(B_1,\ldots,B_{2K})$ is equivalent to the operator being one-to-one (Rudin (1973, pp.99-103)). Suppose that $B_1[v,\mathcal{W}]=0$, $\ldots$, and $B_{2K}[v,\mathcal{W}]=0$. It is easy to see from (7.4) that the derivative of $\mathcal{P}[v^T\mathcal{L}_\theta(\theta_0,\mathcal{A}_0)+\sum_{k=1}^K\mathcal{L}_k(\theta_0,\mathcal{A}_0)[\int h_k d\Lambda_{0k}]]$ along the path $(\theta_0+\epsilon v,\mathcal{A}_0+\epsilon\int\mathcal{W}d\mathcal{A}_0)$ is zero. That is, the information along this path is zero, or $v^T\mathcal{L}_\theta(\theta_0,\mathcal{A}_0)+\sum_{k=1}^K\mathcal{L}_k(\theta_0,\mathcal{A}_0)[\int h_k d\Lambda_{0k}]=0$ almost surely. By (C7), $v=0$ and $\mathcal{W}=0$, so that $(B_1, B_{21},\ldots,B_{2K})$ is one-to-one and invertible.

It follows from (7.1) that, for any $(v,\mathcal{W})\in\mathcal{V}\times\mathcal{Q}^K$,

$$
\begin{aligned}
&n^{1/2}\Big\{v^T(\widehat{\theta}-\theta_0)+\sum_{k=1}^K\int_0^\tau h_k(t)d(\widehat{\Lambda}_k(t)-\Lambda_{0k}(t))\Big\}\\
&=-n^{1/2}(\mathcal{P}_n-\mathcal{P})\Bigg[\widetilde{v}^T\dot{\mathcal{L}}_\theta(\theta_0,\mathcal{A}_0)+\sum_{k=1}^K\dot{\mathcal{L}}_k(\theta_0,\mathcal{A}_0)\Big[\int\widetilde{h}_k d\Lambda_{0k}\Big]\Bigg]\\
&\quad+o\Bigg(n^{1/2}|\widehat{\theta}-\theta_0|+n^{1/2}\sum_{k=1}^K\|\widehat{\Lambda}_k-\Lambda_{0k}\|_{V[0,\tau]}\Bigg),
\end{aligned}
$$

where $(\widetilde{v},\widetilde{h}_1,\ldots,\widetilde{h}_K)=(B_1,B_{21},\ldots,B_{2K})^{-1}(v,h_1,\ldots,h_K)$. Since

$$
|\widehat{\theta}-\theta_0|+\sum_{k=1}^K\|\widehat{\Lambda}_k-\Lambda_{0k}\|_{V[0,\tau]}
$$

$$= \sup_{(v,h_1,\ldots,h_K)\in\mathcal{V}\times\mathcal{Q}^K} \left| v^T(\widehat{\theta}-\theta_0) + \sum_{k=1}^K \int_0^\tau h_k(t)d(\widehat{\Lambda}_k(t)-\Lambda_{0k}(t)) \right|,$$

$$n^{1/2}\left\{|\widehat{\theta}-\theta_0| + \sum_{k=1}^K \|\widehat{\Lambda}_k - \Lambda_{0k}\|_{V[0,\tau]}\right\}$$
$$= O_p(1) + o\left(n^{1/2}|\widehat{\theta}-\theta_0| + n^{1/2}\sum_{k=1}^K \|\widehat{\Lambda}_k - \Lambda_{0k}\|_{V[0,\tau]}\right).$$

Thus, $n^{1/2}\{|\widehat{\theta}-\theta_0| + \sum_{k=1}^K \|\widehat{\Lambda}_k - \Lambda_{0k}\|_{V[0,\tau]}\} = O_p(1)$. Consequently,

$$n^{1/2}\left\{v^T(\widehat{\theta}-\theta_0) + \sum_{k=1}^K \int_0^\tau h_k(t)d(\widehat{\Lambda}_k(t)-\Lambda_{0k}(t))\right\}$$
$$= -n^{1/2}(\mathcal{P}_n - \mathcal{P})\left[\widetilde{v}^T\dot{\mathcal{L}}_\theta(\theta_0,\mathcal{A}_0) + \sum_{k=1}^K \dot{\mathcal{L}}_k(\theta_0,\mathcal{A}_0)\left[\int \widetilde{h}_k d\Lambda_{0k}\right]\right] + o_p(1).$$

We have proved that $n^{1/2}(\widehat{\theta}-\theta_0, \widehat{\mathcal{A}}-\mathcal{A}_0)$ converges weakly to a Gaussian process in $l^\infty(\mathcal{V}\times\mathcal{Q}^K)$. By choosing $h_k = 0$ for $k = 1,\ldots,K$, we see that $v^T\widehat{\theta}$ is an asymptotically linear estimator of $v^T\theta_0$ with influence function $\widetilde{v}^T\dot{\mathcal{L}}_\theta(\theta_0,\mathcal{A}_0) + \sum_{k=1}^K \dot{\mathcal{L}}_k(\theta_0,\mathcal{A}_0)[\int \widetilde{h}_k d\Lambda_{0k}]$. Since the influence function lies in the space spanned by the score functions, $\widehat{\theta}$ is an efficient estimator for θ_0.

It remains to verify that $\eta_{0k}(t;\theta_0,\mathcal{A}_0) < 0$. Under (C6), $\mathcal{P}[\dot{\Psi}_k(\mathcal{O}_i;\theta_0,\mathcal{A}_0)[H_k] /\Psi(\mathcal{O}_i;\theta_0,\mathcal{A}_0)] = \int_0^\tau \eta_{0k}(s;\theta_0,\mathcal{A}_0)dH_k(s)$. The choice of $H_k(s) = I(s \geq t)$ yields $\mathcal{P}[\dot{\Psi}_k(\mathcal{O}_i;\theta_0,\mathcal{A}_0)[I(\cdot \geq t)]/\Psi(\mathcal{O}_i;\theta_0,\mathcal{A}_0)] = \eta_{0k}(t;\theta_0,\mathcal{A}_0)$. On the other hand, the score function along the path $\Lambda_{0k} + \epsilon I(\cdot \geq t)$, with the other parameters fixed at their true values, has zero expectation. We expand this expectation to obtain

$$\mathcal{P}\left[\frac{\dot{\Psi}_k(\mathcal{O}_i;\theta_0,\mathcal{A}_0)[I(\cdot \geq t)]}{\Psi(\mathcal{O}_i;\theta_0,\mathcal{A}_0)}\right] = -\frac{\lambda_k^{-1}(t)dE[I(R_{ik\cdot}(t)>0)N_{ik\cdot}^*(t)]}{dt} < 0.$$

Thus, $\eta_{0k}(t;\theta_0,\mathcal{A}_0) < 0$.

8. Information Matrix

Theorem 2 implies that the functional parameter $\mathcal{A}$ can be estimated at the same rate as the Euclidean parameter θ. Thus, we may treat (3.1) as a parametric log-likelihood with θ and the jump sizes of Λ_k, $k = 1,\ldots,K$, at the observed failure times as the parameters and estimate the asymptotic covariance matrix of the NPMLEs for these parameters by inverting the information matrix. This result is formally stated in Theorem 3. We impose an additional assumption.

(C8) There exists a neighborhood of $(\theta_0, \mathcal{A}_0)$ such that for $(\theta, \mathcal{A})$ in this neighborhood, the first and second derivatives of $\log \Psi(\mathcal{O}_i; \theta, \mathcal{A})$ with respect to θ and along the path $\Lambda_k + \epsilon H_k$ with respect to ϵ satisfy the inequality in (C4).

For any $v \in \mathcal{V}$ and $h_1, \ldots, h_K \in \mathcal{Q}$, we consider the vector $(v^T, \vec{h}_1^T, \ldots, \vec{h}_K^T)^T$, where $\vec{h}_k$ is the vector consisting of the values of $h_k(\cdot)$ at the observed failure times. Let $\mathcal{I}_n$ be the negative Hessian matrix of (3.1) with respect to $\widehat{\theta}$ and the jump sizes of $(\widehat{\Lambda}_1, \ldots, \widehat{\Lambda}_K)$.

Theorem 3. *Assume* (C1)−(C8)*. Then $\mathcal{I}_n$ is invertible for large n, and*

$$\sup_{v\in\mathcal{V}, h_1,\ldots,h_K\in\mathcal{Q}} \left| n\Big(v^T, \vec{h}_1^T, \ldots, \vec{h}_K^T\Big)\mathcal{I}_n^{-1}(v^T, \vec{h}_1^T, \ldots, \vec{h}_K^T)^T - AVar\Big[n^{1/2}\Big\{v^T(\widehat{\theta}-\theta_0) + \sum_{k=1}^K \int h_k d(\widehat{\Lambda}_k - \Lambda_{0k})\Big\}\Big]\right| \to 0$$

in probability, where AVar denotes the asymptotic variance.

Proof. The proof is similar to that of Theorem 3 in Parner (1998); see also Van der Vaart (1998, pp.419-424). First, (7.4) implies that, for any $v \in \mathcal{V}$ and $h_1, \ldots, h_K \in \mathcal{Q}$,

$$-\mathcal{P}\left(\begin{pmatrix} \ddot{\mathcal{L}}_{\theta\theta} & \ddot{\mathcal{L}}_{\theta 1} & \ldots & \ddot{\mathcal{L}}_{\theta K} \\ \vdots & \vdots & \ddots & \vdots \\ \ddot{\mathcal{L}}_{K\theta} & \mathcal{L}_{K1} & \ldots & \mathcal{L}_{KK} \end{pmatrix}\left[\begin{pmatrix} v \\ \int h_1 d\Lambda_{01} \\ \vdots \\ \int h_K d\Lambda_{0K}\end{pmatrix}, \begin{pmatrix} v \\ \int h_1 d\Lambda_{01} \\ \vdots \\ \int h_K d\Lambda_{0K}\end{pmatrix}\right]\right)$$
$$= v^T B_1(v, h_1, \ldots, h_K) + \sum_{k=1}^K \int B_{2k}(v, h_1, \ldots, h_K) h_k d\Lambda_{0k}, \tag{8.1}$$

where $\ddot{\mathcal{L}}$ pertains to the second-order derivative of the log-likelihood function.

On the right-hand side of (7.4), we replace $\mathcal{P}$ by $\mathcal{P}_n$ to obtain two new linear operators B_{n1} and B_{n2k}. It is easy to show that B_{n1} and B_{n2k} converge uniformly to B_1 and B_{2k}, respectively. Under (C8), the results of Lemma 1 apply to the second-order derivatives $\ddot{\mathcal{L}}$ and the operators $(B_1, B_{21}, \ldots, B_{2K})$. By replacing θ_0, Λ_{0k} and $\mathcal{P}$ on both sides of (8.1) with $\widehat{\theta}$, $\widehat{\Lambda}_{0k}$ and $\mathcal{P}_n$, we obtain

$$(v^T, \vec{h}_1^T, \ldots, \vec{h}_K^T)\mathcal{I}_n(v^T, \vec{h}_1^T, \ldots, \vec{h}_K^T)^T$$
$$= v^T B_{n1}(\widetilde{v}, \widetilde{h}_1, \ldots, \widetilde{h}_K) + \sum_{k=1}^K \int B_{n2k}(\widetilde{v}, \widetilde{h}_1, \ldots, \widetilde{h}_K) h_k d\widehat{\Lambda}_k + o_p(1).$$

According to the proof of Theorem 2, $(B_1, B_{21}, \ldots, B_{2K})$ is invertible, and so is $(B_{n1}, \ldots, B_{n2k})$ for large n. Note that $v^T B_{n1}(\widetilde{v}, \widetilde{h}_1, \ldots, \widetilde{h}_K) + \sum_{k=1}^K \int B_{n2k}(\widetilde{v}, \widetilde{h}_1,$

$\ldots, \widetilde{h}_K) h_k d\widehat{\Lambda}_k$ can be written as $(v^T, \vec{h}_1^T, \ldots, \vec{h}_K^T) \times \mathcal{B}_n (v^T, \vec{h}_1^T, \ldots, \vec{h}_K^T)^T$ for some matrix $\mathcal{B}_n$. Therefore $\mathcal{B}_n$ is invertible, and so is $\mathcal{I}_n$. Furthermore,

$$\sup_{v \in \mathcal{V}, h_1, \ldots, h_K \in \mathcal{Q}} \Big| (v^T, \vec{h}_1^T, \ldots, \vec{h}_K^T) \mathcal{I}_n (v^T, \vec{h}_1^T, \ldots, \vec{h}_K^T)^T - (v^T, \vec{h}_1^T, \ldots, \vec{h}_K^T) \mathcal{B}_n (v^T, \vec{h}_1^T, \ldots, \vec{h}_k^T)^T \Big| \to 0.$$

According to Theorem 2, the asymptotic variance of $n^{1/2}\{v^T(\widehat{\theta} - \theta_0) + \sum_{k=1}^K \int h_k d(\widehat{\Lambda}_k - \Lambda_{0k})\}$ is

$$\mathcal{P}\left[\left\{\dot{\mathcal{L}}_\theta^T \widetilde{v} + \sum_{k=1}^K \dot{\mathcal{L}}_k \left[\int \widetilde{h}_k d\Lambda_{0k}\right]\right\}^2\right]$$
$$= -\mathcal{P}\left\{\begin{pmatrix} \ddot{\mathcal{L}}_{\theta\theta} & \ddot{\mathcal{L}}_{\theta 1} & \ldots & \ddot{\mathcal{L}}_{\theta K} \\ \vdots & \vdots & \ddots & \vdots \\ \ddot{\mathcal{L}}_{K\theta} & \mathcal{L}_{K1} & \ldots & \mathcal{L}_{KK} \end{pmatrix} \left[\begin{pmatrix} \widetilde{v} \\ \int \widetilde{h}_1 d\Lambda_{01} \\ \vdots \\ \int \widetilde{h}_K d\Lambda_{0K} \end{pmatrix}, \begin{pmatrix} \widetilde{v} \\ \int \widetilde{h}_1 d\Lambda_{01} \\ \vdots \\ \int \widetilde{h}_K d\Lambda_{0K} \end{pmatrix}\right]\right\},$$

where $(\widetilde{v}, \widetilde{h}_1, \ldots, \widetilde{h}_K)$ is $(B_1, B_{21}, \ldots, B_{2K})^{-1}(v, h_1, \ldots, h_K)$, which can be approximated by $(B_{n1}, B_{n21}, \ldots, B_{n2K})^{-1}(v, h_1, \ldots, h_K)$. Hence, the asymptotic variance can be approximated uniformly in v and h_k's by its empirical counterpart $(v^T, \vec{h}_1^T, \ldots, \vec{h}_K^T)\mathcal{B}_n^{-1}\mathcal{I}_n\mathcal{B}_n^{-1}(\widetilde{v}^T, \vec{\widetilde{h}}_1^T, \ldots, \vec{\widetilde{h}}_K^T)^T$, which is further approximately by $(v^T, \vec{h}_1^T, \ldots, \vec{h}_K^T)\mathcal{I}_n^{-1}(v^T, \vec{h}_1^T, \ldots, \vec{h}_K^T)^T$.

9. Profile Likelihood

Theorem 4. *Let $pl_n(\theta)$ be the profile log-likelihood function for θ, and assume* (C1)$-$(C8). *For any $\epsilon_n = O_p(n^{-1/2})$ and any vector v,*

$$-\frac{pl_n(\widehat{\theta} + \epsilon_n v) - 2pl_n(\widehat{\theta}) + pl_n(\widehat{\theta} - \epsilon_n v)}{n\epsilon_n^2} \to_p v^T \Sigma^{-1} v,$$

where Σ is the limiting covariance matrix of $n^{1/2}(\widehat{\theta} - \theta_0)$. Furthermore, $2\{pl_n(\widehat{\theta}) - pl_n(\theta_0)\} \to_d \chi_d^2$.

Proof. We appeal to Theorem 1 of Murphy and van der Vaart (2000). Specifically, we construct the least favorable submodel for θ_0 and verify all the conditions in their Theorem 1. For notational simplicity, we assume that $K = 1$. It is straightforward to extend to $K > 1$.

It follows from the proof of Theorem 2 that

$$\int_0^\tau B_2(0, h)h^* d\Lambda_0 = -E\left[\ddot{\mathcal{L}}_{\Lambda\Lambda}\left[\int h^* d\Lambda_0, \int h d\Lambda_0\right]\right],$$

where B_2 stands for the operator $(B_{21}, \ldots, B_{2K})$, and $\ddot{\mathcal{L}}_{\Lambda\Lambda}[H_1, H_2]$ denotes the second-order derivative of $\mathcal{L}(\theta, A)$ with respect to Λ along the bi-directions H_1 and H_2. On the other hand,

$$E\left[\dot{\mathcal{L}}_\Lambda\left[\int h^* d\Lambda_0\right]\dot{\mathcal{L}}_\theta\right] = -\int_0^\tau h^*(s)\dot{\mathcal{L}}_\Lambda^*\dot{\mathcal{L}}_\theta d\Lambda_0(s),$$

where $\mathcal{L}_\Lambda^*$ is the dual operator of $\mathcal{L}_\Lambda$ in $L_2[0, \tau]$. Thus, if we choose h such that $B_2(0, h) = -\dot{\mathcal{L}}_\Lambda^*\dot{\mathcal{L}}_\theta$, then

$$E\left[\dot{\mathcal{L}}_\Lambda\left[\int h^* d\Lambda_0\right]\dot{\mathcal{L}}_\theta\right] = -E\left[\ddot{\mathcal{L}}_{\Lambda\Lambda}\left[\int h^* d\Lambda_0, \int h d\Lambda_0\right]\right].$$

By definition, $\int h d\Lambda_0$ is the least favorable direction for θ_0 and $\dot{\mathcal{L}}_\theta - \dot{\mathcal{L}}_\Lambda[\int h d\Lambda_0]$ is the efficient score function. Such an h exists since $B_2(0, \cdot)$ is invertible. In addition, $h \in BV[0, \tau]$. Hence, we can construct the least favorable submodel at (θ, Λ) by $\epsilon \mapsto (\epsilon, \Lambda_\epsilon)$ with $d\Lambda_\epsilon(\theta, \Lambda) = \{1 + (\epsilon - \theta) \cdot h\}d\Lambda$. Clearly, $\Lambda_\theta(\theta, \Lambda) = \Lambda$ and

$$\frac{\partial \mathcal{L}(\epsilon, \Lambda_\epsilon)}{\partial \epsilon}\Big|_{\epsilon=\theta_0, \theta=\theta_0, \Lambda=\Lambda_0} = \dot{\mathcal{L}}_\theta - \dot{\mathcal{L}}_\Lambda\left[\int h d\Lambda_0\right].$$

If $\widetilde{\theta} \to_p \theta_0$ and $\widehat{\Lambda}_{\widetilde{\theta}}$ maximizes the objective function with $\widehat{\theta}$ replaced by $\widetilde{\theta}$, we can use the arguments in the proof of Theorem 1 to show that $\widehat{\Lambda}_{\widetilde{\theta}}$ is consistent. In the likelihood equation for $\widehat{\Lambda}_{\widetilde{\theta}}$, we can use the arguments for the linearization of (7.1) to show that, uniformly in $h \in \mathcal{Q}$,

$$\begin{aligned} & o_p(1) + n^{1/2}(\mathcal{P}_n - \mathcal{P})\left[\dot{\mathcal{L}}_\Lambda(\theta_0, \Lambda_0)\left[\int h d\Lambda_0\right]\right] \\ & = -n^{1/2}\int_0^\tau B_2(0, h) d(\widehat{\Lambda}_{\widetilde{\theta}} - \Lambda_0) + O_p(n^{1/2}|\widetilde{\theta} - \theta_0|) + o_p(n^{1/2}\|\widehat{\Lambda}_{\widetilde{\theta}} - \Lambda_0\|_{V[0,\tau]}). \end{aligned}$$

The arguments for proving the invertibility of (B_1, B_2) show that $h \mapsto B_2(0, h)$ is invertible. Thus,

$$\|\widehat{\Lambda}_{\widetilde{\theta}} - \Lambda_0\|_{V[0,\tau]} = O_p\Big(|\widetilde{\theta} - \theta_0| + n^{-1/2}\Big).$$

By condition (C6), we obtain the no-bias condition, i.e.,

$$E\left[\frac{\partial \mathcal{L}(\epsilon, \Lambda_\epsilon)}{\partial \epsilon}\Big|_{\epsilon=\theta_0, \theta=\widetilde{\theta}, \Lambda=\widehat{\Lambda}_{\widetilde{\theta}}}\right] = O_p\Big(|\widetilde{\theta} - \theta_0| + n^{-1/2}\Big).$$

We have verified conditions (8)−(11) of Murphy and van der Vaart (2000).

Condition (C4), together with Lemma 1, implies that the class

$$\left\{\frac{\partial \mathcal{L}(\epsilon, \Lambda_\epsilon)}{\partial \epsilon} : |\epsilon - \theta_0| < \delta_0, (\theta, \Lambda) \in \mathcal{N}_0\right\}$$

is P-Donsker and that the functions in the class are continuous at (θ_0, Λ_0) almost surely, while condition (C8) implies that the class

$$\left\{ \frac{\partial^2 \mathcal{L}(\epsilon, \Lambda_\epsilon)}{\partial \epsilon^2} : |\epsilon - \theta_0| < \delta_0, (\theta, \Lambda) \in \mathcal{N}_0 \right\}$$

is P-Glivenko-Cantelli and is bounded in $L_2(P)$. Therefore, all the conditions in Murphy and van der Vaart (2000) hold, so that the desired results follows from their Theorem 1.

10. Applications

In this section, we apply the general results to the problems described in Section 2. We identify a set of conditions for each problem under which regularity conditions (C1)−(C8) are satisfied so that the desired asymptotic properties hold. These applications not only provide the theoretical justifications for the work of Zeng and Lin (2007), but also illustrate how the general theory can be applied to specific problems.

10.1. Transformation models with random effects for dependent failure times

We assume the following.

(D1) The parameter value $(\beta_0^T, \gamma_0^T)^T$ belongs to the interior of a compact set Θ in R^d, and $\Lambda_{0k}'(t) > 0$ for all $t \in [0, \tau]$, $k = 1, \ldots, K$.

(D2) With probability one, $Z_{ikl}(\cdot)$ and $\widetilde{Z}_{ikl}(\cdot)$ are in $BV[0, \tau]$ and are left-continuous with bounded left- and right-derivatives in $[0, \tau]$.

(D3) With probability one, $P(C_{ikl} \geq \tau | Z_{ikl}) > \delta_0 > 0$ for some constant δ_0.

(D4) With probability one, n_{ik} is bounded by some integer n_0. In addition, $E[N_{ik\cdot}(\tau)] < \infty$.

(D5) For $k = 1, \ldots, K$, $G_k(x)$ is four-times differentiable such that $G_k(0) = 0, G_k'(x) > 0$, and for any integer $m \geq 0$ and any sequence $0 < x_1 < \ldots < x_m \leq y$,

$$\prod_{l=1}^{m} \{(1 + x_l) G_k'(x_l)\} \exp\{-G_k(y)\} \leq \mu_{0k}^m (1 + y)^{-\kappa_{0k}}$$

for some constants μ_{0k} and $\kappa_{0k} > 0$. In addition, there exists a constant ρ_{0k} such that

$$\sup_x \left\{ \frac{|G_k''(x)| + |G^{(3)}(x)| + |G^{(4)}(x)|}{G'(x)(1 + x)^{\rho_{0k}}} \right\} < \infty.$$

(D6) For any constant $a_1 > 0$,

$$\sup_{\gamma} E\bigg[\int_b \exp\{a_1(N^*_{ik\cdot}(\tau)+1)|b|\} f(b;\gamma)db\bigg] < \infty,$$

and there exists a constant $a_2 > 0$ such that for any γ,

$$\left|\frac{\dot{f}_\gamma(b;\gamma)}{f(b;\gamma)}\right| + \left|\frac{\ddot{f}_\gamma(b;\gamma)}{f(b;\gamma)}\right| + \left|\frac{f^{(3)}_\gamma(b;\gamma)}{f(b;\gamma)}\right| \le O(1)\exp\{a_2(1+|b|)\}.$$

(D7) Consider two types of events: $k \in \mathcal{K}_1$ indicates that event k is recurrent and $k \in \mathcal{K}_2$ indicates that event k is survival time. For $k \in \mathcal{K}_1 \cup \mathcal{K}_2$, if there exist $c_k(t)$ and v such that with probability 1, $c_k(t) + v^T Z_{ikl}(t) = 0$ for $k \in \mathcal{K}_1$ and $c_k(0) + v^T Z_{ikl}(0) = 0$ for $k \in \mathcal{K}_2$, then $v = 0$.

(D8) If there exist constants α_k and α_{0k} such that for any subset $L_k \subset \{1, \ldots, n_{ik}\}$ and for any ω_{kl} and t_{kl},

$$\begin{aligned}
&\int_b \prod_{k\in\mathcal{K}_1}\prod_{l=1}^{n_{ik}} \exp\{\mathrm{i}\omega_{kl} b^T \widetilde{Z}_{ikl}(t_{kl})\} \prod_{k\in\mathcal{K}_2}\prod_{l\in L_k} \exp\{\alpha_k + b^T\widetilde{Z}_{ikl}(0)\} f(b;\boldsymbol{\gamma})db \\
&\quad = \int_b \prod_{k\in\mathcal{K}_1}\prod_{l=1}^{n_{ik}} \exp\{\mathrm{i}\omega_{kl} b^T \widetilde{Z}_{ikl}(t_{kl})\} \prod_{k\in\mathcal{K}_2}\prod_{l\in L_k} \exp\{\alpha_{0k} + b^T\widetilde{Z}_{ikl}(0)\} f(b;\boldsymbol{\gamma}_0)db,
\end{aligned}$$

then $\gamma = \gamma_0$. In addition, if for $k \in \mathcal{K}_2$ and for any t,

$$\begin{aligned}
&\int_b \exp\bigg\{-G_k\bigg(\int_0^t e^{b^T\widetilde{Z}_{ikl}(s)} d\Lambda_1(s)\bigg)\bigg\} f(b;\boldsymbol{\gamma}_0)db \\
&\quad = \int_b \exp\bigg\{-G_k\bigg(\int_0^t e^{b^T\widetilde{Z}_{ikl}(s)} d\Lambda_2(s)\bigg)\bigg\} f(b;\boldsymbol{\gamma}_0)db,
\end{aligned}$$

then $\Lambda_1 = \Lambda_2$. Furthermore, if for some vector v and constant α_k,

$$\begin{aligned}
I(k\in\mathcal{K}_1)&\int_b e^{2b^T\widetilde{Z}_{ikl}(0)} f'(b;\gamma_0)^T v db + I(k\in\mathcal{K}_2)\int_b e^{b^T\widetilde{Z}_{ikl}(0)}(\alpha_k f(b;\gamma_0) \\
&-f'(b;\gamma_0)^T v)db = 0,
\end{aligned}$$

then $v = 0$.

(D1)−(D4) are standard conditions for this type of problem. We show that (D5) holds for all commonly used transformations. We first consider the class of logarithmic transformations $G(x) = \rho\log(1+rx)$ $(\rho > 0, r > 0)$. Clearly,

$$\prod_{k=1}^m \Big\{(1+x_k)G'(y)\Big\}\exp\{-G(y)\} \le \prod_{k=1}^m \bigg\{\frac{\rho r(1+x_k)}{1+rx_k}\bigg\}(1+ry)^{-\rho}$$

$$\leq \left\{\rho r\Big(1+\frac{1}{r}\Big)\right\}^m \min(1,r)^{-\rho}(1+y)^{-\rho}.$$

Thus, in (D5), we can set μ_0 to $\rho r(1+1/r)\min(1,r)^{-\rho}$ and κ_0 to ρ. We can verify the polynomial bounds for $G''(x)/G(x)$, $G^{(3)}(x)/G(x)$ and $G^{(4)}(x)/G(x)$ by direct calculations. We next consider the class of Box-Cox transformations $G(x) = \{(1+x)^\rho - 1\}/\rho$. Clearly,

$$\begin{aligned}
&\prod_{k=1}^m \Big\{(1+x_k)G'(x_k)\Big\}\exp\{-G(y)\} \\
&\quad\leq \prod_{k=1}^m (1+x_k)^\rho \exp\left[-\frac{(1+y)^\rho - 1}{\rho}\right] \\
&\quad\leq (1+y)^{m\rho}\exp\left\{-\frac{(1+y)^\rho}{2\rho}\right\}\exp\left\{-\frac{(1+y)^\rho}{2\rho}\right\}\exp\left(\frac{1}{\rho}\right) \\
&\quad\leq \left\{4\rho + \exp\left(\frac{1}{\rho}\right)\right\}^m (1+y)^{-\rho}.
\end{aligned}$$

Thus, we can set μ_0 to $4\rho + \exp(1/\rho)$ and κ_0 to ρ. The polynomial bounds for $G''(x)/G(x)$, $G^{(3)}(x)/G(x)$ and $G^{(4)}(x)/G(x)$ hold naturally. Finally, we consider the linear transformation model: $H(T) = \beta^T Z + \epsilon$, where ϵ is standard normal. In this case, $G(x) = -\log\{1-\Phi(\log x)\}$, where Φ is the standard normal distribution function. We claim that there exists a constant $\nu_0 > 0$ such that $\phi(x) \leq \nu_0\{1-\Phi(x)\}(1+|x|)$. If $x < 0$, then $\phi(x) \leq (2\pi)^{-1/2} \leq 2(2\pi)^{-1/2}\{1-\Phi(x)\}(1+|x|)$. If $x \geq 0$,

$$\lim_{x\to 0}\frac{\phi(x)}{\{1-\Phi(x)\}(1+x)} = 2(2\pi)^{-1/2}.$$

By the L'Hospital rule,

$$\lim_{x\to\infty}\frac{1-\Phi(x)}{\phi(x)} = \lim_{x\to\infty}\frac{\phi(x)}{\phi(x)x} = 0,$$

$$\begin{aligned}
\lim_{x\to\infty}\frac{\phi(x)}{\{1-\Phi(x)\}(1+x)} &= \lim_{x\to\infty}\frac{-\phi(x)x}{-\phi(x)(1+x) + \{1-\Phi(x)\}} \\
&= \lim_{x\to\infty}\frac{1}{(1+x)/x - \{1-\Phi(x)\}/x\phi(x)} = 1.
\end{aligned}$$

Therefore, $\phi(x)/[\{1-\Phi(x)\}(1+x)]$ is bounded for $x \geq 0$. Without loss of generality, assume that $y > 1$. Clearly,

$$\prod_{k=1}^m \Big\{(1+x_k)G'(x_k)\Big\}\exp\{-G(y)\} = \prod_{k=1}^m \left\{\frac{(1+x_k)\phi(\log(x_k))/x_k}{1-\Phi(\log(x_k))}\right\}\Big\{1-\Phi(\log y)\Big\}.$$

Since $(1+x)\phi(\log(x))/[x\{1-\Phi(\log x)\}]$ is bounded when x is close to zero and it is bounded by a multiplier of $(1+\log x)$ when x is close to ∞, $(1+x)\phi(\log(x))/x\{1-\Phi(\log x)\} \leq \nu_{01} + \nu_{02}\log(1+x)$ for two constants ν_{01} and ν_{02}. Therefore,

$$\prod_{k=1}^{m}\Big\{(1+x_k)G'(x_k)\Big\}\exp\{-G(y)\} \leq \Big\{\nu_{01} + \nu_{02}\log(1+y)\Big\}^m\{1-\Phi(\log y)\}.$$

Since $1-\Phi(x) \leq 2^{1/2}\exp(-x^2/4)$ when $x > 0$, the above expression is bounded by

$$\begin{aligned}
&2^{1/2}\Big\{\nu_{01} + \nu_{02}\log(1+y)\Big\}^m\exp\{-(\log y)^2/4\}\\
&\leq \nu_{03}\Big\{\nu_{01} + \nu_{02}\log(1+y)\Big\}^m\exp\{-\nu_{04}(\log(1+y))^2\}\\
&\leq \nu_{05}^m(1+y)^{-\nu_{04}/2},
\end{aligned}$$

where all the ν's are positive constants. The polynomial bounds for $G''(x)/G(x)$, $G^{(3)}(x)/G(x)$ and $G^{(4)}(x)/G(x)$ follow from the fact that $\phi(x)/\{1-\Phi(x)\} \leq O(1+|x|)$.

Condition (D6) pertains to the tail property of the density function for the random effects $f(b;\gamma)$. For survival data, $N_{ik\cdot}^*(\tau) \leq 1$, so that the first half of condition (D6) is tantamount to that the moment generating function of b exists everywhere. This condition holds naturally when b has a compact support or a Gaussian density tail. The second half of condition (D6) clearly holds for Gaussian density functions.

(D7) and (D8) are sufficient conditions to ensure parameter identifiability and non-singularity of the Fisher information matrix. In most applications, these conditions are tantamount to the linear independence of covariates and the unique parametrization of the random-effects distribution. Specifically, if $\widetilde{Z}_{ikl}$ is time-independent, then the second condition in (D8) is not necessary; if $\widetilde{Z}_{ikl}$ does not depend on k and l, and b has a normal distribution, then the other two conditions in (D8) hold as well provided that $\widetilde{Z}_{ikl}$ is linearly independent with positive probability; if $\widetilde{Z}_{ikl}$ is time-independent and $\mathcal{K}_1$ is non-empty (i.e., at least one event is recurrent), then (D8) can be replaced by the linear independence of $\widetilde{Z}_{ikl}$ for some $k \in \mathcal{K}_1$ and the unique parametrization of $f(b;\gamma)$.

We wish to show that (D1)−(D8) imply (C1)−(C8), so that the desired asymptotic properties hold. Conditions (C1) and (C2) follow naturally from (D1)−(D4). To verify (C3), we note that

$$\Psi(\mathcal{O}_i;\theta,\mathcal{A}) = \int_b \prod_{k=1}^{K}\prod_{l=1}^{n_{ik}}\Omega_{ikl}(b;\beta,\Lambda_k)f(b;\gamma)db,$$

where

$$\Omega_{ikl}(b;\beta,\Lambda_k) = \prod_{t\le\tau}\Big\{R_{ikl}(t)e^{\beta^T Z_{ikl}(t)+b^T\widetilde{Z}_{ikl}(t)}G_k'(q_{ikl}(t))\Big\}^{dN^*_{ikl}(t)} \times \exp\Big\{-G_k(q_{ikl}(\tau))\Big\},$$

and $q_{ikl}(t) = \int_0^t R_{ikl}(s)\exp\{\beta^T Z_{ikl}(s) + b^T \widetilde{Z}_{ikl}(s)\}d\Lambda_k(s)$.

If $\|\Lambda_k\|_{V[0,\tau]}$ are bounded, then $\Omega_{ikl}(b;\beta,\Lambda_k) \ge \exp\{O(1)N^*_{ikl}(\tau)\}I(|b|\le B_0)$ for any fixed constant B_0 such that $P(|b| \le B_0) > 0$. Thus, $\Psi(\mathcal{O}_i;\theta,\mathcal{A})$ is bounded from below by $\exp\{O(1)N^*_{ikl}(\tau)\}$, so that the second half of (C3) holds. It follows from (D5) that

$$\Omega_{ikl}(b;\beta,\Lambda_k) \le O(1)\prod_{t\le\tau}\Big\{R_{ikl}(t)e^{b^T\widetilde{Z}_{ikl}(t)}\Big\}^{dN^*_{ikl}(t)} \times\mu_{0k}^{N^*_{ikl}(\tau)}\prod_{t\le\tau}\Big\{1+q_{ikl}(t)\Big\}^{-dN^*_{ikl}(t)}\Big\{1+q_{ikl}(\tau)\Big\}^{-\kappa_{0k}}.$$

Since $\exp\{\beta^T Z_{ikl}(s) + b^T\widetilde{Z}_{ikl}(s)\} \ge \exp\{-O(1+|b|)\}$, we have $1+q_{ikl}(t) \ge e^{-O(1+|b|)}\{1+\int_0^t R_{ik\cdot}(s)d\Lambda_k(s)\}$, so that

$$\Omega_{ikl}(b;\beta,\Lambda_k) \le O(1)\mu_{0k}^{N^*_{ikl}(\tau)}e^{O(1+N^*_{ikl}(\tau))|b|}\prod_{t\le\tau}\Big\{1+\int_0^t R_{ik\cdot}(s)d\Lambda_k(s)\Big\}^{-dN^*_{ikl}(t)} \times\Big\{1+\int_0^\tau R_{ikl}(s)d\Lambda_k(s)\Big\}^{-\kappa_{0k}}.$$

Thus, the first half of (C3) holds as well.

We now verify (C4). Under (D5),

$$\Big|\Omega_{ikl}(b;\beta,\Lambda_k)\Big| \le \exp\Big\{O(1+N^*_{ikl}(\tau))|b|\Big\},$$

$$\begin{aligned}&\Big|\frac{\partial}{\partial\beta}\Omega_{ikl}(b;\beta,\Lambda_k)\Big|\\ &= \Bigg|\Omega_{ikl}(b;\beta,\Lambda_k)\Bigg[\Bigg\{\int R_{ikl}(t)Z_{ikl}(t)dN^*_{ikl}(t)\\ &\quad+\int R_{ikl}(t)\frac{G_k''(q_{ikl}(t))\int_0^t R_{ikl}(s)e^{\beta^T Z_{ikl}(s)+b^T\widetilde{Z}_{ikl}(s)}Z_{ikl}(s)d\Lambda_k(s)}{G_k'(q_{ikl}(t))}dN^*_{ikl}(t)\Bigg\}\\ &\quad-G_k'(q_{ikl}(\tau))\Bigg\{\int_0^\tau R_{ikl}(s)e^{\beta^T Z_{ikl}(s)+b^T\widetilde{Z}_{ikl}(s)}Z_{ikl}(s)d\Lambda_k(s)\Bigg\}\Bigg]\Bigg|\end{aligned}$$

$$\leq \exp\Big\{O(1+N^*_{ikl}(\tau))(1+|b|)\Big\},$$

$$\begin{aligned}
&\left|\frac{\partial}{\partial\Lambda_k}\Omega_{ikl}(b;\beta,\Lambda_k)[H_k]\right| \\
&\quad = \Bigg|\Omega_{ikl}(b;\beta,\Lambda_k) \\
&\qquad \times\Bigg[\Bigg\{\int R_{ikl}(t)\frac{G''_k(q_{ikl}(t))\int_0^t R_{ikl}(s)e^{\beta^T Z_{ikl}(s)+b^T\widetilde{Z}_{ikl}(s)}dH_k(s)}{G'_k(q_{ikl}(t))}dN^*_{ikl}(t)\Bigg\} \\
&\qquad -G'_k(q_{ikl}(\tau))\Bigg\{\int_0^\tau R_{ikl}(s)e^{\beta^T Z_{ikl}(s)+b^T\widetilde{Z}_{ikl}(s)}dH_k(s)\Bigg\}\Bigg]\Bigg| \\
&\quad \leq \exp\Big\{O(1+N^*_{ikl}(\tau))(1+|b|)\Big\}.
\end{aligned}$$

Thus, it follows from the Mean-Value Theorem that

$$\begin{aligned}
\left|\Omega_{ikl}(b;\beta^{(1)},\Lambda_k)-\Omega_{ikl}(b;\beta^{(2)},\Lambda_k)\right| &= \left|\frac{\partial}{\partial\beta}\Omega_{ikl}(b;\beta^*,\Lambda_k)\right|\left|\beta^{(1)}-\beta^{(2)}\right| \\
&\leq \exp\Big\{O(1+N^*_{ikl}(\tau))|b|\Big\}\left|\beta^{(1)}-\beta^{(2)}\right|,
\end{aligned}$$

$$\begin{aligned}
&\Big|\Omega_{ikl}(b;\beta,\Lambda_k^{(1)})-\Omega_{ikl}(b;\beta,\Lambda_k^{(2)})\Big| \\
&\quad = \left|\frac{\partial}{\partial\Lambda_k}\Omega_{ikl}(b;\beta,\Lambda_k^*)\Big[\Lambda_k^{(1)}-\Lambda_k^{(2)}\Big]\right| \\
&\quad \leq \exp\big\{O(1+N^*_{ikl}(\tau))|b|\Big\} \\
&\qquad \times\Bigg\{\int R_{ikl}(t)\Bigg|\int_0^t e^{\beta^{*T}Z_{ikl}(s)+b^T\widetilde{Z}_{ikl}(s)}d(\Lambda_k^{(1)}-\Lambda_k^{(2)})(s)\Bigg|dN^*_{ikl}(t) \\
&\qquad +\Bigg|\int_0^\tau R_{ikl}(t)e^{\beta^{*T}Z_{ikl}(s)+b^T\widetilde{Z}_{ikl}(s)}d(\Lambda_k^{(1)}-\Lambda_k^{(2)})(s)\Bigg|\Bigg\} \\
&\quad \leq \exp\Big\{O(1+N^*_{ikl}(\tau))(1+|b|)\Big\} \\
&\qquad \times\Bigg\{\int R_{ikl}(t)|\Lambda_k^{(1)}(t)-\Lambda_k^{(2)}(t)|dN^*_{ikl}(t)+\int_0^\tau|\Lambda_k^{(1)}(s)-\Lambda_k^{(2)}(s)|ds\Bigg\},
\end{aligned}$$

where the last inequality follows from integration by parts and the fact that $Z_{ikl}(t)$ and $\widetilde{Z}_{ikl}(t)$ have bounded variations. It then follows from (D6) that $|\Psi(\mathcal{O}_i;\theta^{(1)},\mathcal{A}^{(1)})-\Psi(\mathcal{O}_i;\theta^{(2)},\mathcal{A}^{(2)})|$ is bounded by the right-hand side of the inequality in (C4). By the same arguments, we can verify the bounds for the other three terms in (C4).

To verify (C6), we calculate that

$$\begin{aligned}
&\eta_{0k}(s;\theta,\mathcal{A})\\
&\quad = E\Bigg[\int_b \frac{\prod_{m=1}^K \prod_{l=1}^{n_{im}} \Omega_{iml}(b;\beta,\Lambda_m) f(b;\gamma)}{\int_b \prod_{m=1}^K \prod_{l=1}^{n_{im}} \Omega_{iml}(b;\beta,\Lambda_m) f(b;\gamma) db}\\
&\qquad \times \bigg\{\int_{t\geq s} \frac{G_k''(q_{ikl}(t))}{G_k'(q_{ikl}(t))} dN_{ikl}^*(t) - G_k'(q_{ikl}(\tau))\bigg\} R_{ikl}(s) e^{\beta^T Z_{ikl}(s) + b^T \widetilde{Z}_{ikl}(s)} db\Bigg].
\end{aligned}$$

For $(\theta,\mathcal{A})$ in a neighborhood of $(\theta_0,\mathcal{A}_0)$,

$$\begin{aligned}
&\bigg|\eta_{0k}(s;\theta,\mathcal{A}) - \eta_{0k}(s;\theta_0,\mathcal{A}_0) - \frac{\partial}{\partial\theta}\eta_{0k}(s;\theta_0,\mathcal{A}_0)^T(\theta-\theta_0)\\
&\quad - \sum_{m=1}^K \frac{\partial \eta_{0k}}{\partial \Lambda_m}(s;\theta_0,\mathcal{A}_0)[\Lambda_m - \Lambda_{0m}]\bigg| = o\bigg(|\theta-\theta_0| + \sum_{m=1}^K \|\Lambda_m - \Lambda_{0m}\|_{V[0,\tau]}\bigg).
\end{aligned}$$

Thus, for the second equation in (C6), $\eta_{0km}(s,t;\theta_0,\mathcal{A}_0)$ is obtained from the derivative of η_{0k} with respect to Λ_m along the direction $\Lambda_m - \Lambda_{0m}$, and $\eta_{0k\theta}$ is the derivative of η_{0k} with respect to θ. Likewise, we can obtain the first equation in (C6). It is straightforward to verify the Lipschitz continuity of η_{0km}.

The verification of (C8) is similar to that of (C4), relying on the explicit expressions of $\ddot{\Psi}_{\theta\theta}(\mathcal{O}_i;\theta,\mathcal{A})$ and the first and second derivatives of $\Psi(\mathcal{O}_i;\theta,\mathcal{A}_0 + \epsilon\mathcal{H})$ with respect to ϵ.

It remains to verify the two identifiability conditions under (D7) and (D8). To verify (C5), suppose that $(\beta,\gamma,\Lambda_1,\ldots,\Lambda_k)$ yields the same likelihood as $(\beta_0,\gamma_0,\Lambda_{10},\ldots,\Lambda_{k0})$. That is,

$$\begin{aligned}
&\int_b \prod_{k=1}^K \prod_{l=1}^{n_{ik}} \lambda_k(t)^{dN_{ikl}^*(t)} \Omega_{ikl}(b;\beta,\Lambda_k) f(b;\gamma) db\\
&\quad = \int_b \prod_{k=1}^K \prod_{l=1}^{n_{ik}} \lambda_{k0}(t)^{dN_{ikl}^*(t)} \Omega_{ikl}(b;\beta_0,\Lambda_{k0}) f(b;\gamma_0) db.
\end{aligned}$$

We perform the following operations on both sides sequentially for $k = 1,\ldots,K$ and $l = 1,\ldots,n_{ik}$.

(a) If the kth type of event pertains to survival time, for the lth subject of this type of event, the first equation is obtained with $R_{ikl}(t) = 1$ and $dN_{ikl}^*(t) = 0$ for any $t \leq \tau$, i.e., the subject does not experience any event in $[0,\tau]$. The second equation is obtained by integrating t from t_{kl} to τ on both sides under the scenario that $R_{ikl}(t) = 1$ and $N_{ikl}^*(t)$ has a jump at t, i.e, the subject experiences the event at time t_{kl}. We then take the difference between these two equations. In the resulting equation, the terms $\lambda_k(t)^{dN^*_{ikl}(t)}\Omega_{ikl}(b;\beta,\Lambda_k)$ and $\lambda_{k0}(t)^{dN^*_{ikl}(t)}\Omega_{ikl}(b;\beta_0,\Lambda_{k0})$ are replaced by $\exp\{-G_k(\int_0^{t_{kl}}\exp\{\beta^T Z_{ikl}(s)+b^T\widetilde{Z}_{ikl}(s)\}d\Lambda_k)\}$ and $\exp\{-G_k(\int_0^{t_{kl}}\exp\{\beta_0^T Z_{ikl}(s)+b^T\widetilde{Z}_{ikl}(s)\}d\Lambda_{k0})\}$, respectively.

(b) If the kth type of event is recurrent, for the lth subject of this type of event, we let $R_{ikl}(t)=1$ and let $N^*_{ikl}(t)$ have jumps at $s_1,\ldots,s_m$ and $s'_1,\ldots,s'_{m'}$ for any arbitrary $(m+m')$ times in $[0,\tau]$. We integrate $s_1,\ldots,s_m$ from 0 to t_{kl} and integrate $s'_1,\ldots,s'_{m'}$ from 0 to τ. In the obtained equation, $\lambda_k(t)^{dN^*_{ikl}(t)}\Omega_{ikl}(b;\beta,\Lambda_k)$ is replaced by $\{G_k(q_{ikl}(t_{kl}))\}^m\{G_k(q_{ikl}(\tau))\}^{m'}$ on both sides. Note that m and m' are arbitrary. We then multiple both sides by $\{(\mathrm{i}\omega_{kl})^m/m!\}/m'!$ and sum over $m,m'=0,1,\ldots$. On both sides of the resulting equation, the terms associated with k and l are replaced by $\exp\{\mathrm{i}\omega_{kl}G_k(q_{ikl}(t_{kl}))\}$.

After these sequential operations, we obtain

$$\begin{aligned}&\int_b\prod_{k\in\mathcal{K}_1}\prod_{l=1}^{n_{ik}}\exp\Big\{\mathrm{i}\omega_{kl}G_k(q_{ikl}(t_{kl}))\Big\}\\&\qquad\times\prod_{k\in\mathcal{K}_2}\prod_{l=1}^{n_{ik}}\exp\Big\{-G_k(q_{ikl}(t_{kl}))\Big\}f(b;\boldsymbol{\gamma})db\\&\quad=\int_b\prod_{k\in\mathcal{K}_1}\prod_{l=1}^{n_{ik}}\exp\Big\{\mathrm{i}\omega_{kl}G_k(q_{ikl0}(t_{kl}))\Big\}\\&\qquad\times\prod_{k\in\mathcal{K}_2}\prod_{l=1}^{n_{ik}}\exp\Big\{-G_k(q_{ikl0}(t_{kl}))\Big\}f(b;\boldsymbol{\gamma}_0)db.\end{aligned}$$

For survival time, we can let any subject from the n_{ik} subjects have $t_{kl}=0$, which results in

$$\begin{aligned}&\int_b\prod_{k\in\mathcal{K}_1}\prod_{l=1}^{n_{ik}}\exp\Big\{\mathrm{i}\omega_{kl}G_k(q_{ikl}(t_{kl}))\Big\}\\&\qquad\times\prod_{k\in\mathcal{K}_2}\prod_{l=1}^{n_{ik}}\left[\frac{1}{\xi_{kl}}+\exp\Big\{-G_k(q_{ikl}(t_{kl}))\Big\}\right]f(b;\boldsymbol{\gamma})db\\&\quad=\int_b\prod_{k\in\mathcal{K}_1}\prod_{l=1}^{n_{ik}}\exp\Big\{\mathrm{i}\omega_{kl}G_k(q_{ikl0}(t_{kl}))\Big\}\\&\qquad\times\prod_{k\in\mathcal{K}_2}\prod_{l=1}^{n_{ik}}\left[\frac{1}{\xi_{kl}}+\exp\Big\{-G_k(q_{ikl0}(t_{kl}))\Big\}\right]f(b;\boldsymbol{\gamma}_0)db,\end{aligned}$$

where ξ_{kl} is any positive variable.

The above expression implies that $\{G_k(q_{ikl}(t)), k \in \mathcal{K}_1\}$ as a function of

$$b_1 \sim \prod_{k\in\mathcal{K}_2}\prod_{l=1}^{n_{ik}}\left[\frac{1}{\xi_{kl}} + \exp\Big\{-G_k(q_{ikl}(t_{kl}))\Big\}\right] f(b;\gamma)$$

has the same distribution as $\{G_k(q_{ikl0}(t)), k \in \mathcal{K}_1\}$ as a function of

$$b_2 \sim \prod_{k\in\mathcal{K}_2}\prod_{l=1}^{n_{ik}}\left[\frac{1}{\xi_{kl}} + \exp\Big\{-G_k(q_{ikl0}(t_{kl}))\Big\}\right] f(b;\gamma_0);$$

so this is true between $\{q_{ikl}(t)\}$ and $\{q_{ikl0}(t)\}$ because of the one-to-one mapping. Thus, the distributions of $\{\log q'_{ikl}(t)\}$ and $\{\log q'_{ikl0}(t)\}$ should also agree and they have the same expectation. Now let $t_{kl} = 0$ for $k \in \mathcal{K}_2$. Since $E[b_1] = E[b_2] = 0$, we obtain $\log \lambda_k(t) + \beta^T Z_{ikl}(t) = \log \lambda_{k0}(t) + \beta_0^T Z_{ikl}(t)$ for $k \in \mathcal{K}_1$. The above arguments also yield

$$\begin{aligned}
&\int_b \prod_{k\in\mathcal{K}_1}\prod_{l=1}^{n_{ik}} \exp\Big\{b^T \widetilde{Z}_{ikl}(t_{kl})\Big\} \\
&\qquad\times \prod_{k\in\mathcal{K}_2}\prod_{l=1}^{n_{ik}}\left[\frac{1}{\xi_{kl}} + \exp\Big\{-G_k(q_{ikl}(t_{kl}))\Big\}\right] f(b;\gamma)db \\
&\quad= \int_b \prod_{k\in\mathcal{K}_1}\prod_{l=1}^{n_{ik}} \exp\Big\{b^T \widetilde{Z}_{ikl}(t_{kl})\Big\} \\
&\qquad\times \prod_{k\in\mathcal{K}_2}\prod_{l=1}^{n_{ik}}\left[\frac{1}{\xi_{kl}} + \exp\Big\{-G_k(q_{ikl0}(t_{kl}))\Big\}\right] f(b;\gamma_0)db.
\end{aligned}$$

We compare the coefficients of ξ_{kl} for $k \in \mathcal{K}_2$. This yields that for any subset $L_k \subset \{1, \ldots, n_{ik}\}$,

$$\begin{aligned}
&\int_b \prod_{k\in\mathcal{K}_1}\prod_{l=1}^{n_{ik}} \exp\Big\{\mathrm{i}\omega_{kl} b^T \widetilde{Z}_{ikl}(t_{kl})\Big\} \\
&\qquad\times \prod_{k\in\mathcal{K}_2}\prod_{l\in L_k} \exp\Big\{-G_k(q_{ikl}(t))\Big\} f(b;\gamma)db \\
&\quad= \int_b \prod_{k\in\mathcal{K}_1}\prod_{l=1}^{n_{ik}} \exp\Big\{\mathrm{i}\omega_{kl} b^T \widetilde{Z}_{ikl}(t_{kl})\Big\} \\
&\qquad\times \prod_{k\in\mathcal{K}_2}\prod_{l\in L_k} \exp\Big\{-G_k(q_{ikl0}(t))\Big\} f(b;\gamma_0)db.
\end{aligned}$$

We differentiate the above expression with respect to t_{kl} at 0 for $k \in \mathcal{K}_2$. It then follows from (D8) that $\log \lambda_k(0) - \log \lambda_{0k}(0) + (\beta - \beta_0)^T Z_{ikl}(0) = 0$ and

$\gamma = \gamma_0$. Thus, (D7) implies that $\beta = \beta_0$ and $\lambda_k(t) = \lambda_{0k}(t)$ for $k \in \mathcal{K}_1$. On the other hand, for any fixed $k \in \mathcal{K}_2$, we let $t_{k'l'} = 0$ if $k' \neq k$ or $l' \neq l$. Thus, $\int_b \exp\{-G_k(q_{ikl}(t_{kl}))\} f(b; \gamma_0) db = \int_b \exp\{-G_k(q_{0ikl}(t_{kl}))\} f(b; \gamma_0) db$. Therefore, $\Lambda_k = \Lambda_{0k}$ for $k \in \mathcal{K}_2$ according to (D8).

To verify (C7), we write $v = (v_\beta, v_\gamma)$. We perform operations (a) and (b) on the score equation in (C7). The arguments used in proving the identifiability yield

$$\begin{aligned} \int_b \Bigg[\sum_{k \in \mathcal{K}_1} \sum_{l=1}^{n_{ik}} \mathrm{i}\omega_{kl} A_{ikl}(t_{kl}) G_k(q_{ikl0}(t_{kl})) - \sum_{k \in \mathcal{K}_2} \sum_{l \in L_k} A_{ikl}(t_{kl}) \\ + \frac{f'(b; \gamma_0)^T v_\gamma}{f(b; \gamma_0)} \Bigg] \exp\Bigg\{ \sum_{k \in \mathcal{K}_1} \sum_{l=1}^{n_{ikl}} \mathrm{i}\omega_{kl} G_k(q_{ikl0}(t_{kl})) \\ - \sum_{k \in \mathcal{K}_2} \sum_{l \in L_k} G_k(q_{ikl0}(t_{kl})) \Bigg\} f(b; \gamma_0) db = 0, \end{aligned} \tag{10.1}$$

where $A_{ikl}(t) = \int_0^t (h_k(s) + Z_{ikl}(s)^T v_\beta) e^{\beta_0^T Z_{ikl}(s) + b^T \widetilde{Z}_{ikl}(s)} d\Lambda_{k0}(s) G'_k(q_{ikl0}(t))$. We differentiate (10.1) with respect to t_{kl} twice at 0 for $k \in \mathcal{K}_1$. Comparison of the coefficients for ω_{kl} yields $\int_b e^{2b^T \widetilde{Z}_{ikl}(0)} f'(b; \gamma_0)^T v_\gamma db = 0$. We also differentiate (10.1) with respect to t_{kl} at 0 for $k \in \mathcal{K}_2$. Thus, for each $k \in \mathcal{K}_2$ and $l = 1, \ldots, n_{ik}$, $\int_b (h_k(0) + Z_{ikl}(0)^T v_\beta) e^{b^T \widetilde{Z}_{ikl}(0)} f(b; \gamma_0) db = -G'_k(0) \int_b e^{b^T \widetilde{Z}_{ikl}(0)} f'(b; \gamma_0)^T v_\gamma db$. It then follows from (D8) that $v_\gamma = 0$. For fixed k_0 and l_0, with the fact of $v_\gamma = 0$, the score equation under operations (a) and (b), where in (a) we let $dN^*_{ikl}(t) = 0$ for any $t \leq \tau$ and in (b) we let $m = 0$ whenever $k \neq k_0$ or $l \neq l_0$, becomes a homogeneous integral equation for $h_{k_0}(t) + Z_{ik_0l_0}(t)^T v_\beta$. The equation has a trivial solution, so $h_{k_0}(t) + Z_{ik_0l_0}(t)^T v_\beta = 0$. Since k_0 and l_0 are arbitrary, (D7) implies that $h_k = 0$ and $v_\beta = 0$.

Remark 2. For survival time, (D5) is required to hold only for $m = 0$ and $m = 1$.

Remark 3. The above results do not apply directly to the proportional hazards model with gamma frailty because (D6) does not hold when b has a gamma distribution. It is mathematically convenient to handle this model because the marginal hazard function has an explicit form. The likelihood is a special case of ours with

$$\begin{aligned} \Psi(\mathcal{O}_i; \theta, \Lambda) = \prod_{j=1}^{n_i} \prod_{t \leq \tau} Y_{ij}(t; \beta)^{dN_{ij}(t)} \prod_{t \leq \tau} \Big\{ 1 + \theta N_{i\cdot}(u-) \Big\}^{dN_{i\cdot}(t)} \\ \times \Big\{ 1 + \theta \int_0^\tau Y_{i\cdot}(u; \beta) d\Lambda(u) \Big\}^{-(1/\theta + N_{i\cdot}(\tau))} \end{aligned}$$

in the notation of Parner (1998). Clearly, Ψ satisfies (C3) when $\theta > 0$. The other conditions can be verified in the same manner as before.

Remark 4. Our theory does not cover the case in which the true parameter values lie on the boundary of Θ. It is delicate to deal with the boundary problem. One possible solution is to follow the idea of Parner (1998) by extending the definition of the likelihood function outside Θ and verifying (C2)−(C8) for the extended likelihood function.

Remark 5. We have assumed known transformations. We may allow G_k to belong to a parametric family of distributions, say $G_k(\cdot;\psi)$, where ψ is a parameter in a compact set. Then θ contains ψ. Our results and proofs apply to this situation if (D5) holds uniformly in ψ and the two identifiability conditions are satisfied.

10.2. Joint models for repeated measures and failure times

For the (parametric) generalized linear mixed model, the likelihood can be viewed as a special case of that of Section 10.1 except that there is an additional parameter α in $f(y|x;b)$. We assume that (D1)−(D8) hold but with (D6) replaced by the following.

(D6') For any constant $a_1 > 0$,

$$\sup_{\alpha,\gamma} E\left[\int_b \exp\left\{a_1(N_i^*(\tau)+1)|b|\right\}\prod_{j=1}^{n_i} f(Y_{ij}|X_{ij};b)f(b;\gamma)db\right] < \infty,$$

and there exists a constant $a_2 > 0$ such that for any γ and α,

$$\sum_{k=1}^{3}\left|\frac{f_\alpha^{(k)}(Y_{ij}|X_{ij},b)}{f(Y_{ij}|X_{ij},b)}\right| + \left|\frac{f_\gamma^{(k)}(b;\gamma)}{f(b;\gamma)}\right| \le r_3(\mathcal{O}_i)\exp\left\{a_2(1+|b|)\right\}$$

almost surely, where $r_3(\mathcal{O}_i)$ is a random variable in $L_2(P)$.

Under these conditions, the desired asymptotic properties follow from the arguments of Section 10.1.

Under the semiparametric linear transformation model for continuous repeated measures, the likelihood is in the form of that of Section 2.2 with $K = 2$ and $n_{i2} = n_i$, where the time to the second type of failure is defined by Y_{ij} (assuming without loss of generality that $Y_{ij} \ge 0$). Thus, if we regard Y_{ij} as a right-censored observation when it is greater than a very large value (i.e., the upper limit of detection), then the asymptotic results given in Section 10.1 hold.

When such an upper limit does not exist, the estimator for $\widetilde{\Lambda}$ can be unbounded when sample size goes to infinity. Then our proof of Theorem 1 does not apply.

10.3. Transformation models for counting processes

We verify (C1)$-$(C8) under the following conditions.

(E1) The parameter value $(\beta_0^T, \gamma_0^T)^T$ belongs to the interior of a compact set Θ in R^d, and $\Lambda_0'(t) > 0$ for all $t \in [0, \tau]$.

(E2) With probability one, $P(C \geq \tau | Z) > \delta_0 > 0$ for some constant δ_0.

(E3) Condition (D5) holds.

(E4) With probability one, $Z(\cdot)$ and $\widetilde{Z}$ are in $BV[0, \tau]$ and are left-continuous with bounded left- and right-derivatives in $[0, \tau]$.

(E5) If $\gamma^T \widetilde{Z}$ is equal to a constant with probability one, then $\gamma = 0$. In addition, if $\beta^T Z(t) = c(t)$ for a deterministic function $c(t)$ with probability one, then $\beta = 0$.

In this case,

$$\begin{aligned}\Psi(\mathcal{O}_i; \theta, \Lambda) = \prod_{t \leq \tau} & \left(R_i(t) e^{\beta^T Z_i(t) + \gamma^T \widetilde{Z}_i} \left\{ 1 + \int_0^t R_i(s) e^{\beta^T Z_i(s)} d\Lambda(s) \right\}^{e^{\gamma^T \widetilde{Z}_i} - 1} \right. \\ & \left. \times G' \left[\left\{ 1 + \int_0^t R_i(s) e^{\beta^T Z_i(s)} d\Lambda(s) \right\}^{e^{\gamma^T \widetilde{Z}_i}} \right] \right)^{dN_i^*(t)} \\ & \times \exp\left(- G\left[\left\{ 1 + \int_0^\tau R_i(s) e^{\beta^T Z_i(s)} d\Lambda(s) \right\}^{e^{\gamma^T \widetilde{Z}_i}} \right] \right).\end{aligned}$$

By (D5),

$$\begin{aligned}\prod_{t \leq \tau} & \left(R_i(t) e^{\beta^T Z_i(t) + \gamma^T \widetilde{Z}_i} \left\{ 1 + \int_0^t R_i(s) e^{\beta^T Z_i(s)} d\Lambda(s) \right\}^{e^{\gamma^T \widetilde{Z}_i} - 1} \right. \\ & \left. \times G' \left[\left\{ 1 + \int_0^t R_i(s) e^{\beta^T Z_i(s)} d\Lambda(s) \right\}^{e^{\gamma^T \widetilde{Z}_i}} \right] \right)^{dN_i^*(t)} \\ & \times \exp\left(- G\left[\left\{ 1 + \int_0^\tau R_i(s) e^{\beta^T Z_i(s)} d\Lambda(s) \right\}^{e^{\gamma^T \widetilde{Z}_i}} \right] \right) \\ & \leq \mu_1^{N_i^*(\tau)} \prod_{t \leq \tau} \left\{ 1 + \int_0^t R_i(s) e^{\beta^T Z_i(s)} d\Lambda(s) \right\}^{-dN_i^*(t)} \\ & \times \left\{ 1 + \int_0^\tau R_i(s) e^{\beta^T Z_i(s)} d\Lambda(s) \right\}^{-\kappa e^{\gamma^T \widetilde{Z}_i}}\end{aligned}$$

for some constant μ_1. Thus, (C3) follows from the boundedness of $\gamma^T \widetilde{Z}_i$. We can verify the other conditions by using the arguments of Section 10.1.

To verify the first identifiability condition, we assume that $N_i^*(t)$ has jumps at $x, x_1, \ldots, x_m$ for some integer m. After integrating both sides of the equation in (C5) over $x_1, \ldots, x_m$ from 0 to τ and integrating x from x to τ, we obtain

$$\begin{aligned}
&\left(G\left[\left\{1+\int_0^\tau e^{\beta_0^T Z_i(t)} d\Lambda_0(t)\right\}^{e^{\gamma_0^T \widetilde{Z}_i}}\right]-G\left[\left\{1+\int_0^x e^{\beta_0^T Z_i(t)} d\Lambda_0(t)\right\}^{e^{\gamma_0^T \widetilde{Z}_i}}\right]\right) \\
&\quad\times\left(G\left[\left\{1+\int_0^\tau e^{\beta_0^T Z_i(t)} d\Lambda_0(t)\right\}^{e^{\gamma_0^T \widetilde{Z}_i}}\right]-G(1)\right)^m \\
&\quad\times\exp\left(-G\left[\left\{1+\int_0^\tau e^{\beta_0^T Z_i(t)} d\Lambda_0(t)\right\}^{e^{\gamma_0^T \widetilde{Z}_i}}\right]+G(1)\right) \\
&=\left(G\left[\left\{1+\int_0^\tau e^{\beta^{*T} Z_i(t)} d\Lambda^*(t)\right\}^{e^{\gamma^{*T} \widetilde{Z}_i}}\right]-G\left[\left\{1+\int_0^x e^{\beta^{*T} Z_i(t)} d\Lambda^*(t)\right\}^{e^{\gamma^{*T} \widetilde{Z}_i}}\right]\right) \\
&\quad\times\left(G\left[\left\{1+\int_0^\tau e^{\beta^{*T} Z_i(t)} d\Lambda^*(t)\right\}^{e^{\gamma^{*T} \widetilde{Z}_i}}\right]-G(1)\right)^m \\
&\quad\times\exp\left(-G\left[\left\{1+\int_0^\tau e^{\beta^{*T} Z_i(t)} d\Lambda^*(t)\right\}^{e^{\gamma^{*T} \widetilde{Z}_i}}\right]+G(1)\right).
\end{aligned}$$

Multiplying both sides of this equation by $1/m!$ and summing over $m \geq 0$, we obtain

$$\begin{aligned}
&G\left[\left\{1+\int_0^\tau e^{\beta_0^T Z_i(t)} d\Lambda_0(t)\right\}^{e^{\gamma_0^T \widetilde{Z}_i}}\right]-G\left[\left\{1+\int_0^x e^{\beta_0^T Z_i(t)} d\Lambda_0(t)\right\}^{e^{\gamma_0^T \widetilde{Z}_i}}\right] \\
&\quad=G\left[\left\{1+\int_0^\tau e^{\beta^{*T} Z_i(t)} d\Lambda^*(t)\right\}^{e^{\gamma^{*T} \widetilde{Z}_i}}\right]-G\left[\left\{1+\int_0^x e^{\beta^{*T} Z_i(t)} d\Lambda^*(t)\right\}^{e^{\gamma^{*T} \widetilde{Z}_i}}\right].
\end{aligned}$$

Setting $N_i^*(\tau)=0$ in the likelihood function yields

$$G\left[\left\{1+\int_0^\tau e^{\beta_0^T Z_i(t)} d\Lambda_0(t)\right\}^{e^{\gamma_0^T \widetilde{Z}_i}}\right]=G\left[\left\{1+\int_0^\tau e^{\beta^{*T} Z_i(t)} d\Lambda^*(t)\right\}^{e^{\gamma^{*T} \widetilde{Z}_i}}\right].$$

Thus

$$\left\{1+\int_0^x e^{\beta_0^T Z_i(t)} d\Lambda_0(t)\right\}^{e^{\gamma_0^T \widetilde{Z}_i}}=\left\{1+\int_0^x e^{\beta^{*T} Z_i(t)} d\Lambda^*(t)\right\}^{e^{\gamma^{*T} \widetilde{Z}_i}}.$$

Then $\Lambda^*(t)$ is absolutely continuous with respect to t. Differentiating both sides with respect to x and letting $x = 0$ yield $\lambda^*(0) > 0$. When x converges to zero, the left-hand side is $[\exp\{\beta_0^T Z_i(0)\}\lambda_0(0)x]^{e^{\gamma_0^T \widetilde{Z}_i}} + o(x^{e^{\gamma_0^T \widetilde{Z}_i}})$ while the right-hand side is $t[\exp\{\beta^{*T} Z_i(0)\}\lambda^*(0)x]^{e^{\gamma^{*T}\widetilde{Z}_i}} + o(x^{e^{\gamma^{*T}\widetilde{Z}_i}})$. Thus, $\gamma_0^T \widetilde{Z}_i = \gamma^{*T}\widetilde{Z}_i$. By (E5), $\gamma_0 = \gamma^*$. Furthermore, $e^{\beta_0^T Z_i(t)} d\Lambda_0(t)/dt = e^{\beta^{*T} Z_i(t)} d\Lambda^*(t)/dt$. It follows from (E5) that $\beta_0 = \beta^*$ and $\Lambda_0 = \Lambda^*$.

To verify (C7), we assume that the score function along $(\beta_0 + \epsilon h_\beta, \gamma_0 + \epsilon h_\gamma, d\Lambda_0 + \epsilon h d\Lambda_0)$ is zero. Equivalently, if we let $g_0(t) = \{1 + \int_0^t e^{\beta_0^T Z_i(s)} d\Lambda_0(s)\}^{\gamma_0^T \widetilde{Z}_i}$, then we obtain

$$
\begin{aligned}
0 = & \int h(t) R_i(t) dN_i^*(t) + \int R_i(t)\Big\{h_\beta^T Z_i(t) + h_\gamma^T \widetilde{Z}_i\Big\} dN_i^*(t) \\
& + \int \frac{R_i(t)(e^{\gamma^T \widetilde{Z}_i} - 1)}{1 + \int_0^t e^{\beta_0^T Z_i(s)} d\Lambda_0(s)} \bigg[\int_0^t e^{\beta_0^T Z_i(s)} \Big\{h_\beta^T Z_i(s) + h(s)\Big\} d\Lambda_0(s)\bigg] dN_i^*(t) \\
& + \int R_i(t) h_\gamma^T \widetilde{Z}_i e^{\gamma^T \widetilde{Z}_i} \log\Big\{1 + \int_0^t e^{\beta_0^T Z_i(s)} d\Lambda_0(s)\Big\} dN_i^*(t) \\
& + \int R_i(t) \frac{G''(g_0(t))}{G'(g_0(t))} g_0(t) h_\gamma^T \widetilde{Z}_i e^{\gamma^T \widetilde{Z}_i} \log\Big\{1 + \int_0^t e^{\beta_0^T Z_i(s)} d\Lambda_0(s)\Big\} dN_i^*(t) \\
& + \int R_i(t) \frac{G''(g_0(t))}{G'(g_0(t))} g_0(t) \bigg[\frac{e^{\gamma_0^T \widetilde{Z}_i} \int_0^t e^{\beta_0^T Z_i(s)} \{h_\beta^T Z_i(s) + h(s)\} d\Lambda_0(s)}{1 + \int_0^t e^{\beta_0^T Z_i(s)} d\Lambda_0(s)}\bigg] dN_i^*(t) \\
& - G'(g_0(\tau)) g_0(\tau) h_\gamma^T \widetilde{Z}_i e^{\gamma^T \widetilde{Z}_i} \log\Big\{1 + \int_0^\tau e^{\beta_0^T Z_i(s)} d\Lambda_0(s)\Big\} \\
& - G'(g_0(\tau)) g_0(\tau) \frac{e^{\gamma_0^T \widetilde{Z}_i}}{1 + \int_0^\tau e^{\beta_0^T Z_i(s)} d\Lambda_0(s)} \int_0^\tau e^{\beta_0^T Z_i(s)} \Big\{h_\beta^T Z_i(s) + h(s)\Big\} d\Lambda_0(s).
\end{aligned}
$$

We multiply both sides by the likelihood function and let $N_i^*(t)$ have jumps at times $t_1, \ldots, t_m$. We integrate t_1 from 0 to t and t_l, $1 < l \le m$ from 0 to τ. By multiplying the resulting equation by $1/(m-k)!$ and summing over $m = 1, \ldots,$ we obtain

$$
h_\gamma^T \widetilde{Z}_i \log\Big\{1 + \int_0^t e^{\beta_0^T Z_i(s)} d\Lambda_0(s)\Big\} + \frac{\int_0^t e^{\beta_0^T Z_i(s)} \{h_\beta^T Z_i(s) + h(s)\} d\Lambda_0(s)}{1 + \int_0^t e^{\beta_0^T Z_i(s)} d\Lambda_0(s)} = 0.
$$

Differentiation with respect to t then yields

$$
h_\gamma^T \widetilde{Z}_i + \Big\{h_\beta^T Z_i(t) + h(t)\Big\} - \frac{\int_0^t e^{\beta_0^T Z_i(s)} \{h_\beta^T Z_i(s) + h(s)\} d\Lambda_0(s)}{1 + \int_0^t e^{\beta_0^T Z_i(s)} d\Lambda_0(s)} = 0.
$$

Combining the above two equations, we have

$$\Big\{h_\beta^T Z_i(t) + h(t)\Big\} - \frac{\int_0^t e^{\beta_0^T Z_i(s)}\{h_\beta^T Z_i(s) + h(s)\}d\Lambda_0(s)}{1 + \int_0^t e^{\beta_0^T Z_i(s)} d\Lambda_0(s)} \times \left[1 + \frac{1}{\log\{1 + \int_0^t e^{\beta_0^T Z_i(s)} d\Lambda_0(s)\}}\right] = 0.$$

This is a homogeneous integral equation for $h_\beta^T Z_i(t) + h(t)$ and has zero solution. That is, $h_\beta^T Z_i(t) + h(t) = 0$. It follows from (E5) that $h(t) = 0$ and $h_\beta = 0$. Thus, $h_\gamma = 0$.

11. Concluding Remarks

We have developed a general asymptotic theory for the NPMLEs with right censored data and shown that this theory applies to the models considered by Zeng and Lin (2007). This theory can also be used to establish the desired asymptotic properties for other existing semiparametric models, particularly the models mentioned in Sections 7.1−7.4 of Zeng and Lin (2007), as well as those that may be invented in the future. It is much simpler to verify the set of sufficient conditions identified in this paper than to prove the asymptotic results from scratch. Conditions (C1) and (C2) are standard conditions required in all censored-data regression; (C3), (C4) and (C6) are certain smoothness conditions that can be verified directly, as demonstrated in Section 10; (C5) and (C7) are two minimal identifiability conditions that need to be verified for any specific problem.

Although the basic structures of our proofs mimic those of Murphy (1994, 1995) and Parner (1998), our technical arguments are innovative and substantially more difficult because we deal with a very general form of likelihood function rather than specific problems. In all previous work, verification of the Donsker property relies on the specific expressions of the functions, whereas our Lemma 1 provides a universal way to verify this property. In verifying the invertibility of the information operator, all previous work requires an explicit expression of the information operator that is identified as the sum of an invertible operator and a compact operator, whereas we allow a very generic form of information operator obtained from the likelihood function (3.1). Murphy and van der Vaart (2001) stated that the consistency of NPMLEs needs to be proved on a case-by-case basis; however, we were able to prove the consistency for a very general likelihood function. Although we borrowed the partitioning idea of Murphy (1994), our technical arguments are very different because of the generic form of the likelihood.

In some applications, the failure times are subject to left truncation in addition to right censoring. To accommodate general censoring/truncation patterns, we define $N(t)$ as the number of events observed by time t and $R(t)$ as the at-risk indicator at time t, reflecting both left truncation and right censoring. Assume that the truncation time has positive mass at time 0, so that (C2) is satisfied. Then all the results continue to hold.

This paper is concerned with the theoretical aspect of the NPMLEs and complements the work of Zeng and Lin (2007). The interested readers are referred to the latter for the calculations of the NPMLEs and for the use of the semiparametric regression models and NPMLEs in practice. The latter also provides rationale for the kind of model considered in Sections 2 and 10 of this paper. Although the latter contains some theoretical elements, this paper presents the theory (especially the regularity conditions) in a more rigorous manner and provides all the proofs.

References

Heitjan, D. F. and Rubin, D. B. (1991). Ignorability and coarse data. *Ann. Statist.* **19**, 2244-2253.

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

Murphy, S. A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23**, 182-198.

Murphy, S. A. and van der Vaart, A. W. (2000). On profile likelihood. *J. Amer. Statist. Assoc.* **95**, 449-485.

Murphy, S. A. and van der Vaart, A. W. (2001). Semiparametric mixtures in case-control studies. *J. Multivariate Anal.* **79** 1-32.

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

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

Van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

Zeng, D. and Lin, D. Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data (with discussion). *J. Roy. Statist. Soc. Ser. B*, **69**, 507-564.

Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599-7420, U.S.A.

E-mail: dzeng@bios.unc.edu

epartment of Biostatistics, University of North Carolina, Chapel Hill, NC 27599-7420, U.S.A.

E-mail: lin@bios.unc.edu

(Received January 2008; accepted January 2009)