Statistica Sinica **17**(2007), 1023-1046

A PROFILE LIKELIHOOD THEORY FOR THE CORRELATED GAMMA-FRAILTY MODEL WITH CURRENT STATUS FAMILY DATA

I-Shou Chang[1], Chi-Chung Wen[2] and Yuh-Jenn Wu[3]

[1]*National Health Research Institutes,* [2]*Tamkang University and* [3]*Chung Yuan Christian University*

Abstract: A profile likelihood inference is made for the regression coefficient and frailty parameters in the correlated gamma-frailty model for current status family data. With the introduction of an identifiability assumption, the identifiability of the parameters and the existence of the nonparametric maximum likelihood estimate (NPMLE) are established, the consistency and convergence rate of the NPMLE are obtained, the invertibility of the efficient Fisher information matrix is proved, and a quadratic expansion of the profile likelihood is established. From these, we show that the NPMLE of the parameters of interest is asymptotically normal and efficient, its covariance matrix can be estimated consistently by means of the profile likelihood, and the likelihood ratio test is asymptotically chi-squared. A simulation study is carried out to illustrate the numerical performance of the likelihood ratio test.

Key words and phrases: Current status family data, least favorable submodel, likelihood ratio statistic, nonparametric maximum likelihood estimate, profile likelihood.

1. Introduction

As explained in Parner (1998) and Yashin, Vaupel and Iachine (1995), an appropriate survival model for the analysis of family data with covariates is the correlated gamma-frailty model, henceforth CGFM. This model extends the Cox's proportional hazards model by the introduction of a frailty variable, which acts multiplicatively on the baseline hazard function, and consists of a component common to every individual in the family and a component specific to each individual. Parner (1998) provided an asymptotic theory for nonparametric maximum likelihood estimation in the CGFM for right censored family data. In this paper, we are interested in the likelihood approach to statistical inference in the CGFM for current status family data.

Let T_{ik}, C_{ik} and Z_{ik} be the survival time, the examination time, and the covariate of the ith individual in the kth family in a study of current status family data. We assume there are $m \geq 2$ members in each family and there are K families. Since every subject in the study is examined at a random observation time

C_{ik}, and at this time it is observed whether the survival time T_{ik} has occurred or not, the observed data is $\{(C_{1k}, \Delta_{1k}, Z_{1k}, \ldots, C_{mk}, \Delta_{mk}, Z_{mk}) \mid k = 1, \ldots, K\}$, where $\Delta_{ik} = I_{[T_{ik} \leq C_{ik}]}$ indicates whether T_{ik} has occurred before C_{ik}. Both T_{ik} and C_{ik} take values in $[0, \infty)$.

We suppress the k's in the notation when considering a single family. Thus $X = (C_1, \Delta_1, Z_1, \ldots, C_m, \Delta_m, Z_m)$ denotes the data from a single family.

Let $H_0, H_1, \ldots, H_m$ be independent gamma distributed random variables with parameters $(\theta_1\theta_\cdot^{-2}, \theta_\cdot^{-1})$, $(\theta_2\theta_\cdot^{-2}, \theta_\cdot^{-1})$, $\ldots$, $(\theta_2\theta_\cdot^{-2}, \theta_\cdot^{-1})$ respectively, where $(\theta_1, \theta_2) \in [0, \infty)^2 \backslash \{(0, 0)\}$, and $\theta_\cdot = \theta_1 + \theta_2$. We assume that, given $Z_1 = z_1, \ldots, Z_m = z_m$ and $H_0 = \eta_0, \ldots, H_m = \eta_m$, the cumulative hazard function of T_i at time t is

$$e^{\theta_3 z_i}(\eta_0 + \eta_i)\Lambda(t), \tag{1.1}$$

where $\Lambda(\cdot)$ is a nondecreasing deterministic baseline function, and θ_3 is a real number.

The quantity $\eta_0 + \eta_i$ in (1.1) is called the frailty for the ith individual, where η_0 is a common component for all individuals in the family and η_i is an individual component. We note that the frailty variable $H_0 + H_i$ is gamma distributed with mean 1 and variance $\theta_\cdot$. The correlation between $H_0 + H_i$ and $H_0 + H_j$, $i \neq j$, is $\theta_1\theta_\cdot^{-1}$. Hence, if θ_2 is zero, the correlation reduces to 1, and (1.1) is the so-called shared gamma frailty model.

We assume further that given $Z_1, \ldots, Z_m$, $H_0, \ldots, H_m$, the variables $T_1, \ldots,$ T_m, $C_1, \ldots, C_m$ are conditionally independent, the random vectors $(C_1, \ldots, C_m,$ $Z_1, \ldots, Z_m)$ and $(H_0, \ldots, H_m)$ are independent, and the joint distribution of $(C_1, \ldots, C_m, Z_1, \ldots, Z_m)$ does not involve $\theta_1, \theta_2, \theta_3$ and Λ. Let $\eta = (\eta_0, \ldots, \eta_m)$ and $\theta = (\theta_1, \theta_2, \theta_3)$. Using (1.1) and the preceding assumptions, we know the likelihood for X given $H_0 = \eta_0, \ldots, H_m = \eta_m$ is proportional to

$$q(\theta_3, \Lambda; \eta, X) = \prod_{i=1}^{m} \left[1 - e^{-e^{\theta_3 Z_i}(\eta_0+\eta_i)\Lambda(C_i)}\right]^{\Delta_i} \left[e^{-e^{\theta_3 Z_i}(\eta_0+\eta_i)\Lambda(C_i)}\right]^{1-\Delta_i}. \tag{1.2}$$

Multiplying (1.2) by the joint density of $(H_0, H_1, \ldots, H_m)$, denoted by $p(\eta; \theta_1, \theta_2)$, and integrating over η, we get the likelihood for X:

$$lik(\theta, \Lambda; X) = \int_{[0,\infty)^{m+1}} p(\eta; \theta_1, \theta_2) q(\theta_3, \Lambda; \eta, X) d\eta. \tag{1.3}$$

We note that if $\gamma(y; a, b) = (b^a/\Gamma(a))y^{a-1}e^{-by}$ denotes the density of the gamma distribution with shape parameter a and scale parameter b, then

$$p(\eta; \theta_1, \theta_2) = \gamma(\eta_0; \theta_1\theta_\cdot^{-2}, \theta_\cdot^{-1}) \prod_{i=1}^{m} \gamma(\eta_i; \theta_2\theta_\cdot^{-2}, \theta_\cdot^{-1}).$$

The parameter space for (θ, Λ) we consider is $\Theta \times \mathcal{L}$, where Θ is a compact subset of $([0,\infty)^2 \backslash \{(0,0)\}) \times \mathcal{R}^1$ and

$$\mathcal{L} = \{\Lambda : [0,\tau) \to [0,\infty) \mid \Lambda(0) = 0,\ \Lambda \text{ is nondecreasing and right continuous}\}.$$

Here $\tau \in (0,\infty]$. Throughout, we assume that the true baseline cumulative hazard function Λ_0 is continuous. In Sections 3, 4 and 5, we suppose the true parameter $\theta_0 = (\theta_{10}, \theta_{20}, \theta_{30})$ is an interior point of Θ. This paper studies the inference problem regarding the frailty parameters θ_1, θ_2, and the regression coefficient θ_3 based on observations of X only.

Statistical inference in Cox's proportional hazards model for current status data was studied by Huang and Wellner (1995), Huang (1996) and Murphy and van der Vaart (1997, 1999, 2000), among others. In particular, they show that the nonparametric maximum likelihood estimate, henceforth NPMLE, of the regression coefficient is asymptotically normal and efficient with $\sqrt{K}$ convergence rate, the likelihood ratio test for the regression parameter is asymptotically chi-squared, and the covariance matrix of the NPMLE can be estimated consistently by means of the profile likelihood. In fact, Murphy and van der Vaart (2000) provide a set of conditions under which semiparametric profile likelihoods admit asymptotic quadratic expansions, and present many of the above results as consequences of the quadratic expansion, consistency of the NPMLE, and the invertibility of the efficient Fisher information matrix.

The purpose of this paper is to establish, with the introduction of an identifiability assumption, the consistency of the NPMLE, the invertibility of the efficient Fisher information, and the asymptotic quadratic expansion for the semiparametric profile likelihood. Based on these results and Murphy and van der Vaart (2000) we obtain an asymptotic profile likelihood theory. We note that Bickel and Ritov (2000) and Murphy and van der Vaart (2000) pointed out that the hardest part of a profile likelihood theory might be the verification of the general conditions described, for example, in Murphy and van der Vaart (2000), and this is borne out here. For example, we need to show that the efficient score is Lipschitz, without the availability of its closed form; we need an upper bound for the entropy of the log-likelihoods; we need to use the identifiability assumption to study the modulus of continuity of certain empirical process indexed by log-likelihoods.

This paper is organized as follows. Section 2 establishes the identifiability of the parameters and the existence and the consistency of the NPMLE under certain regularity conditions. These conditions are reasonable, and can be verified computationally in applications. Section 3 exhibits the efficient score function and the efficient Fisher information matrix, and indicates the invertibility of the latter. Section 4 provides a convergence rate of the NPMLE in an appropriate norm. Section 5 establishes a quadratic expansion of the profile likelihood for θ, and derives from it the asymptotic normality and efficiency of the NPMLE of θ, the asymptotic distribution of the profile likelihood ratio statistic, and a consistent estimate of the covariance matrix of the NPMLE of θ. Section 6 presents a simulation study to indicate the numerical performance of the profile likelihood ratio statistic, and Section 7 discusses computational issues for future studies.

Throughout, let P_0 denote the underlying distribution. For a real vector ν, let ν^T denote its transpose, ν_i its ith component, and $\|\nu\|$ its Euclidean norm. We use the notations $o_P(1)$ and $O_P(1)$, respectively, for a sequence of random vectors converging to zero in probability and being uniformly tight.

We also use the notations P_K and G_K, respectively, for the empirical distribution and the empirical process for the random sample $\{X_1,\ldots,X_K\}$ of X. Moreover, we use the operator notation for evaluating expectation. Thus for every measurable g and probability measure P, we have

$$P_K g = \frac{1}{K}\sum_{k=1}^{K} g(X_k), \qquad Pg = \int g dP,$$

$$G_K g = \sqrt{K}(P_K - P_0)g = \frac{1}{\sqrt{K}}\sum_{k=1}^{K}(g(X_k) - P_0 g).$$

2. Nonparametric Maximum Likelihood Estimate

This section contains three subsections. The first studies the parameter identifiability; the second and the third establish, respectively, the existence and consistency of the NPMLE $(\hat{\theta}_K, \hat{\Lambda}_K)$ of (θ_0, Λ_0). The following assumptions are made.

(A1) Given $(Z_1,\ldots,Z_m) = (z_1,\ldots,z_m)$, each examination variable C_i has a common continuous conditional density function whose support is an interval $[\tau_1,\tau_2]$, with $1/M < \Lambda_0(\tau_1) \le \Lambda_0(\tau_2) < M$ for some constant $M > 0$.

(A2) Each individual covariate Z_i is bounded and non-degenerate.

(A3) (Identifiability) There exists $(c_1^*,\ldots,c_m^*)$ in $(\tau_1,\tau_2)^m$ for which there are $m+3$ different values of $(\delta_1,\ldots,\delta_m,z_1,\ldots,z_m)$ such that if

$$\left(\sum_{i=1}^{3} u_i \frac{\partial}{\partial\theta_i} + \sum_{i=1}^{m} u_{i+3}\frac{\partial}{\partial y_i}\right)\Bigg|_{(\theta,y_1,\ldots,y_m)=(\theta_0,\Lambda_0(c_1^*),\ldots,\Lambda_0(c_m^*))}$$

$$\log\int p(\eta;\theta_1,\theta_2)\prod_{i=1}^{m}\Big[1-e^{-e^{\theta_3 z_i}(\eta_0+\eta_i)y_i}\Big]^{\delta_i}\Big[e^{-e^{\theta_3 z_i}(\eta_0+\eta_i)y_i}\Big]^{1-\delta_i} d\eta = 0 \quad (2.1)$$

for each of these $m+3$ values, then $u_1 = \cdots = u_{m+3} = 0$. Here $\delta_i \in \{0,1\}$ for every $i = 1, \ldots, m$, and $(z_1, \ldots, z_m)$ is in the support of the distribution of $(Z_1, \ldots, Z_m)$.

(A4) Λ_0 is continuously differentiable on $[\tau_1, \tau_2]$ with positive derivative.

Remarks. Assumption **(A3)** is needed in establishing the identifiability of the parameters (see Theorem 2.1), the consistency of the NPMLE, the invertibility of the efficient Fisher information matrix for θ at (θ_0, Λ_0) (see Theorem 3.3), and a convergence rate of the NPMLE (see Section 4). This indicates that in the general framework, Assumption **(A3)** plays the same fundamental role here as the identifiability Assumption II plays in Chang, Hsiung, Wang and Wen (2005) concerning NPMLE in the Cox-gene model. However, we would like to point out that, except for the identifiability of parameters, the proofs of the major theorems in the present paper are markedly different from those in Chang et al. (2005); the main difference comes from the fact that the NPMLE in Chang et al. (2005) can be viewed as a Z-estimator, meaning that it is the zero of estimating functions, and that here it is still an M-estimator, the maximizer of a criterion function. A general discussion of M-estimators and Z-estimators can be found in van der Vaart and Wellner (1996) and van der Vaart (1998).

Without loss of generality, we assume that all the random variables are defined on a sample space Ω with a specific σ-field.

2.1. Identifiability of the parameters

Theorem 2.1. *There exists $d^* > 0$ such that if $\|(\theta - \theta_0, (\Lambda - \Lambda_0)(c_1^*), \ldots, (\Lambda - \Lambda_0)(c_m^*))\| < d^*$, and $lik(\theta, \Lambda) = lik(\theta_0, \Lambda_0)$ a.s. under P_0, then $\theta = \theta_0$ and $\Lambda = \Lambda_0$ on $[\tau_1, \tau_2]$.*

Proof. For every positive integer n, we know from (1.2), **(A1)**, and the conditional independence of $T_1, \ldots, T_m, C_1, \ldots, C_m$ given $Z_1, \ldots, Z_m, H_0, H_1, \ldots, H_m$ that

$$P_0 \begin{pmatrix} C_1 \in [c_1, c_1 + \frac{1}{n}), C_i \in [c_i^*, c_i^* + \frac{1}{n}) \text{ for } i = 2, \ldots, m; \\ \Delta_i = 0, Z_i \in (z_i - \frac{1}{n}, z_i + \frac{1}{n}) \text{ for } i = 1, \ldots, m \end{pmatrix} > 0 \tag{2.2}$$

for every $(c_1, z_1, \ldots, z_m)$ in the support of the distribution of $(C_1, Z_1, \ldots, Z_m)$. This shows that there exists ω_n in Ω such that $Z_i(\omega_n) \in (z_i - 1/n, z_i + 1/n)$, $\Delta_i(\omega_n) = 0$ for every $i = 1, \ldots, m$, $C_1(\omega_n) \in [c_1, c_1 + 1/n)$, $C_i(\omega_n) \in [c_i^*, c_i^* + 1/n)$ for every $i = 2, \ldots, m$, and $lik(\theta, \Lambda; X(\omega_n)) = lik(\theta_0, \Lambda_0; X(\omega_n))$. Letting n go to infinity in $lik(\theta, \Lambda, X(\omega_n)) = lik(\theta_0, \Lambda_0; X(\omega_n))$, we obtain

$$\int p(\eta; \theta_1, \theta_2) \exp(-e^{\theta_3 z_1}(\eta_0 + \eta_1)\Lambda(c_1)) \prod_{i=2}^{m} \exp(-e^{\theta_3 z_i}(\eta_0 + \eta_i)\Lambda(c_i^*)) d\eta$$

$$= \int p(\eta;\theta_{10},\theta_{20})\exp(-e^{\theta_{30}z_1}(\eta_0+\eta_1)\Lambda_0(c_1))\prod_{i=2}^{m}\exp(-e^{\theta_{30}z_i}(\eta_0+\eta_i)\Lambda_0(c_i^*))d\eta. \tag{2.3}$$

Using the fact that both sides of (2.3) are monotone functions in c_1, it suffices to show $\theta=\theta_0$ and $\Lambda(c_i^*)=\Lambda_0(c_i^*)$ for $i=2,\ldots,m$ to establish the identifiability.

Let $\delta_i\in\{0,1\}$ for $i=1,\ldots,m$. Considering $c_1=c_1^*$ and $(\Delta_1,\ldots,\Delta_m)=(\delta_1,\ldots,\delta_m)$ in (2.2), and using the argument leading to (2.3), we get

$$\int p(\eta;\theta_1,\theta_2)\prod_{i=1}^{m}\left[1-e^{-e^{\theta_3 z_i}(\eta_0+\eta_i)\Lambda(c_i^*)}\right]^{\delta_i}\left[e^{-e^{\theta_3 z_i}(\eta_0+\eta_i)\Lambda(c_i^*)}\right]^{1-\delta_i}d\eta$$
$$= \int p(\eta;\theta_1,\theta_2)\prod_{i=1}^{m}\left[1-e^{-e^{\theta_{30} z_i}(\eta_0+\eta_i)\Lambda_0(c_i^*)}\right]^{\delta_i}\left[e^{-e^{\theta_{30} z_i}(\eta_0+\eta_i)\Lambda_0(c_i^*)}\right]^{1-\delta_i}d\eta.$$

Let $G:\mathcal{R}^{m+3}\mapsto\mathcal{R}^{m+3}$ be the vector valued function whose components are of the form

$$(\theta,y_1,\ldots,y_m)\mapsto\log\int p(\eta;\theta_1,\theta_2)\prod_{i=1}^{m}\left[1-e^{-e^{\theta_3 z_i}(\eta_0+\eta_i)y_i}\right]^{\delta_i}\left[e^{-e^{\theta_3 z_i}(\eta_0+\eta_i)y_i}\right]^{1-\delta_i}d\eta.$$

Here $(\delta_1,\ldots,\delta_m,z_1,\ldots,z_m)$ are those $m+3$ different values in **(A3)**. It suffices to show that G is locally invertible in order to obtain the identifiability.

Applying the Inverse Function Theorem (see, for example, Theorem 9.24 in Rudin (1976)), the locally invertibility of G in a neighborhood of $(\theta_0,\Lambda_0(c_1^*),\ldots,\Lambda_0(c_m^*))$ follows if the determinant of the Jacobian of G, denoted by $J_G(\theta,y_1,\ldots,y_m)$, evaluated at $(\theta_0,\Lambda_0(c_1^*),\ldots,\Lambda_0(c_m^*))$ is nonzero. Since J_G is an analytic function, it is zero only on a nowhere dense closed subset of $\mathcal{R}^{m+3}$ if it is not identically zero. Therefore, as long as J_G is not zero at $(\theta_0,\Lambda_0(c_1^*),\ldots,\Lambda_0(c_m^*))$, which is equivalent to **(A3)**, we can find an appropriate d^* such that the conclusion of this theorem is valid. This completes the proof.

Remarks. Although the above proof is similar to the one for Proposition A.1 in Chang et al. (2005), and could have been omitted, we keep it here because its argument appears several times in the rest of this paper. The proof suggests a method to check the identifiability assumption **(A3)**. We now illustrate it by considering the model with true parameters $\theta_0=(1,1,0.5)$ and $\Lambda_0(t)=\log(100/(100-t))$, and $m=3$ members in each family. Let $c_1^*=45, c_2^*=50$ and $c_3^*=55$. Since the determinant of the linear mapping from $\mathcal{R}^6$ to $\mathcal{R}^6$ obtained from (2.1) by specifying $(\delta_1,\delta_2,\delta_3,z_1,z_2,z_3)=(1,1,1,0,0,0)$, $(1,1,0,0,0,0)$, $(1,0,1,0,1,1)$, $(0,1,1,0,0,1)$, $(0,1,0,1,0,0)$ and $(0,0,0,1,0,0)$ is equal to 0.1866, which is not zero, we know **(A3)** is satisfied, and Theorem 2.1 indicates that there is a neighborhood of (θ_0, Λ_0) on which the parameters are identifiable. Because the above determinant is an real analytic function of $(\theta_0, \Lambda_0(c_1^*), \ldots, \Lambda_0(c_m^*))$, its zero set is closed and nowhere dense, and if the determinant is not zero at one point, it is never zero on a neighborhood of it. Therefore, if the identifiability is established for a point in the parameter space, it is also established at each point in a neighborhood of it.

2.2. Existence of NPMLE

Let $X_1, \ldots, X_K$ be i.i.d. copies of X, then, according to (1.3), the likelihood for the data $\{(C_{1k}, \Delta_{1k}, Z_{1k}, \ldots, C_{mk}, \Delta_{mk}, Z_{mk}) \mid k = 1, \ldots, K\}$ is

$$L_K(\theta, \Lambda) = \prod_{k=1}^{K} \int_{[0,\infty)^{m+1}} p(\eta; \theta_1, \theta_2) \prod_{i=1}^{m} \left[1 - e^{-e^{\theta_3 Z_{ik}}(\eta_0+\eta_i)\Lambda(C_{ik})}\right]^{\Delta_{ik}} \times \left[e^{-e^{\theta_3 Z_{ik}}(\eta_0+\eta_i)\Lambda(C_{ik})}\right]^{1-\Delta_{ik}} d\eta. \quad (2.4)$$

Since only the values of Λ at the C_{ik} matter in (2.4), all the estimates $\hat{\Lambda}_K$ of Λ_0 considered in this paper are right continuous nondecreasing step functions with possible jump points C_{ik}.

Theorem 2.2. *If the set $\{(ik, i'k') \mid C_{ik} < C_{i'k'}, \Delta_{ik} = 1, \Delta_{i'k'} = 0\}$ is nonempty, then there exists $(\hat{\theta}_K, \hat{\Lambda}_K)$ that maximizes $L_K(\theta, \Lambda)$ subject to $\theta \in \Theta$ and Λ in the aforesaid and constrained class.*

The condition in Theorem 2.2 is theoretically interesting. Consider, for example, the situation that T_{ik} refers to age of onset of a certain disease. In this case, violation of the condition means all the early examined subjects are not affected and the late examined are affected, which indicates that this age of onset has little variance and hence little statistical study of the problem is needed.

Proof. For an estimate $\hat{\Lambda}_K$ of Λ_0, we take $\hat{y}_{ik} = \hat{\Lambda}_K(C_{ik})$ and $\hat{Y}_K$ to be the matrix $(\hat{y}_{ik})$. Then the maximum likelihood estimate of $\theta_0 = (\theta_{10}, \theta_{20}, \theta_{30})$ and Λ_0 is the $\hat{\theta}_K = (\hat{\theta}_{1K}, \hat{\theta}_{2K}, \hat{\theta}_{3K})$ and $\hat{\Lambda}_K$, represented by $\hat{Y}_K$ that maximizes

$$\psi(\theta, Y) = \prod_{k=1}^{K} \int p(\eta; \theta_1, \theta_2) \prod_{i=1}^{m} \left[1 - e^{-e^{\theta_3 Z_{ik}}(\eta_0+\eta_i)y_{ik}}\right]^{\Delta_{ik}} \left[e^{-e^{\theta_3 Z_{ik}}(\eta_0+\eta_i)y_{ik}}\right]^{1-\Delta_{ik}} d\eta,$$

subject to $(\theta, Y) \in \Theta \times \mathcal{D}$, where

$$\mathcal{D} = \{(y_{ik}) \in R^{m\times K} \mid 0 \leq y_{ik} \leq y_{i'k'} \ if \ C_{ik} \leq C_{i'k'} \text{ for every pair } (ik, i'k')\}.$$

For fixed θ, we first show that there exists an element $\hat{\Lambda}_K(\cdot, \theta)$ in $\mathcal{L}$ that maximizes $L_K(\theta, \Lambda)$ under the above constraint.

Let $C_{(1)} \leq \cdots \leq C_{(mK)}$ denote the order statistic of $\{C_{ik} \mid i = 1, \ldots, m,\ k = 1, \ldots, K\}$, and let $\Delta_{(j)} = \Delta_{ik}$, $Z_{(j)} = Z_{ik}$, $y_{(j)} = \Lambda(C_{ik})$ if $C_{(j)} = C_{ik}$. We note that if $\Delta_{(1)} = 0$, then $\hat{\Lambda}_K(C_{(1)}, \theta) = 0$, and if $\Delta_{(mK)} = 1$, then $\hat{\Lambda}_K(C_{(mK)}, \theta) = \infty$. This shows that the terms associated with $\Delta_{(1)} = 0$ and $\Delta_{(mK)} = 1$ in ψ are 1, and hence do not contribute anything to ψ. Therefore, without loss of generality, we may assume that $\Delta_{(1)} = 1$ and $\Delta_{(mK)} = 0$ in establishing the existence of $\hat{\Lambda}_K(\cdot, \theta)$.

In view of

$$|\psi(\theta, Y)| \leq \int p(\eta; \theta_1, \theta_2) \exp(-e^{\theta_3 Z_{(mK)}}(\eta_0 + \eta_1) y_{(mK)}) d\eta$$
$$= (1 + \theta . e^{\theta_3 Z_{(mK)}} y_{(mK)})^{-\theta_.^{-1}}, \tag{2.5}$$

there exists $d_0 > 0$ such that

$$\max_{y_{(mK)} \leq d_0} \psi(\theta, Y) > \sup_{y_{(mK)} > d_0} \psi(\theta, Y). \tag{2.6}$$

Because ψ is a continuous function, it has a maximizer on any compact set. This together with (2.6) gives the existence of $\hat{\Lambda}_K(\cdot, \theta)$.

Using (2.5), we can show ψ is uniformly continuous on $\Theta \times \mathcal{D}$, and hence the mapping $\theta \mapsto L_K(\theta, \hat{\Lambda}_K(\cdot, \theta))$ is continuous. Since Θ is compact, the maximizer $\hat{\theta}_K$ exists. Let $\hat{\Lambda}_K(\cdot) = \hat{\Lambda}_K(\cdot, \hat{\theta}_K)$. Since $\sup_{(\theta,\Lambda)} L_K(\theta, \Lambda) = \sup_\theta L_K(\theta, \hat{\Lambda}_K(\cdot, \theta))$, we know $(\hat{\theta}_K, \hat{\Lambda}_K)$ maximizes (2.4). This completes the proof.

2.3. Consistency

The following theorem can be established in the framework of Wald that studies the consistency of maximum likelihood estimates (see, for example, van der Vaart (1998), pp.47-51); its proof is hence omitted.

Theorem 2.3. *$\hat{\theta}_K \to \theta_0$ a.s. and $\sup_{t \in [\tau_1, \tau_2]} |\hat{\Lambda}_K(t) - \Lambda_0(t)| \to 0$ a.s..*

3. Efficient Score

The purpose of this section is to find the efficient score and to show the invertibility of the Fisher information matrix. Readers are referred to Bickel, Klaassen, Ritov and Wellner (1993) and van der Vaart (1998) for these definitions and concepts.

Let

$$l_\theta(\theta, \Lambda)(X) = \Big(l_{\theta_1}(\theta, \Lambda)(X), l_{\theta_2}(\theta, \Lambda)(X), l_{\theta_3}(\theta, \Lambda)(X) \Big),$$

where $l_{\theta_i}(\theta, \Lambda)(X) = \frac{\partial}{\partial \theta_i} \log lik(\theta, \Lambda; X)$. $l_{\theta_i}(\theta, \Lambda)$ and $l_\theta(\theta, \Lambda)$ are the score function for θ_i and θ at (θ, Λ), respectively.

Let $\Lambda_\varepsilon = \Lambda + \varepsilon h$, where ε is positive and h is a nondecreasing and nonnegative function defined on $[0, \tau]$. The score function for Λ in the direction h at (θ, Λ), denoted by $l_\Lambda(\theta, \Lambda)[h]$, is defined by

$$l_\Lambda(\theta, \Lambda)[h](X) = \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \log lik(\theta, \Lambda_\varepsilon; X).$$

A little calculation shows that

$$l_\Lambda(\theta, \Lambda)[h](X) = \sum_{i=1}^{m} h(C_i) W_i(\theta, \Lambda; X), \tag{3.1}$$

where

$$W_i(\theta, \Lambda; X) = lik(\theta, \Lambda; X)^{-1} \\ \times \int p(\eta; \theta_1, \theta_2) q(\theta_3, \Lambda; \eta, X) e^{\theta_3 Z_i}(\eta_0 + \eta_i) \left[\frac{\Delta_i}{1 - e^{-e^{\theta_3 Z_i}(\eta_0+\eta_i)\Lambda(C_i)}} - 1 \right] d\eta.$$

We consider the closed linear span (in $L^2(P_0)$) of the score functions $l_\Lambda(\theta, \Lambda)$ $[h]$ for $h \in \mathcal{H}$, the set of all bounded functions defined on $[0, \tau]$ with $\|h\|_{BV} < \infty$. Here the bounded variation norm $\|h\|_{BV}$ is defined to be the sum of the absolute value of $h(\tau_1)$ and the total variation of h on the interval $[\tau_1, \tau_2]$. $\mathcal{H}$ is a Banach space under this norm. Let A be the continuous operator from $\mathcal{H}$ to $\mathcal{H}$ defined by

$$(Ah)(u) = \sum_{i=1}^{m} \sum_{j=1}^{m} E_0(h(C_j) W_j(\theta_0, \Lambda_0; X) W_i(\theta_0, \Lambda_0; X) | C_i = u), \tag{3.2}$$

which is motivated by (3.6) below. Using the following Lemma 3.2, the inverse A^{-1} of A exists. Let $\mathbf{h}^* = (h_1^*, h_2^*, h_3^*)$ be defined by

$$h_j^*(u) = \left(A^{-1} (\sum_{i=1}^{m} E_0 \left[l_{\theta_j}(\theta_0, \Lambda_0; X) W_i(\theta_0, \Lambda_0; X) | C_i = \cdot \ \right]) \right)(u), \tag{3.3}$$

for $j = 1, 2, 3$. In fact, (3.3) becomes

$$h_j^*(u) = \frac{E_0 \left[l_{\theta_j}(\theta_0, \Lambda_0; X) \sum_{i=1}^{m} W_i(\theta_0, \Lambda_0; X) | C_1 = u \ \right]}{E_0 \left[(\sum_{i=1}^{m} W_i(\theta_0, \Lambda_0; X))^2 | C_1 = u \ \right]} \tag{3.4}$$

when all the members in the same family share a common examination time C_1.

We note that the functions $\{h_j^* \mid j = 1, 2, 3\}$ are unique only up to null sets relative to Q, the distribution of examination variable C_i.

Let $l_\Lambda(\theta_0, \Lambda_0)[\mathbf{h}^*] = (l_\Lambda(\theta_0, \Lambda_0)[h_1^*], l_\Lambda(\theta_0, \Lambda_0)[h_2^*] \ , l_\Lambda(\theta_0, \Lambda_0)[h_3^*])$.

Theorem 3.1. *The efficient score function for* θ *at* (θ_0, Λ_0) *is* $\tilde{l}_0 = l_\theta(\theta_0, \Lambda_0) - l_\Lambda(\theta_0, \Lambda_0)[\mathbf{h}^*]$.

We note that the efficient Fisher information matrix for θ at (θ_0, Λ_0), denoted by I_0, is $I_0 = P_0 \tilde{l}_0^T \tilde{l}_0$. We will show that I_0 is positive definite in Theorem 3.3.

Proof. It suffices to show that

$$P_0(l_{\theta_j}(\theta_0, \Lambda_0) - l_\Lambda(\theta_0, \Lambda_0)[h_j^*])(l_\Lambda(\theta_0, \Lambda_0)[h]) = 0, \tag{3.5}$$

for every $h \in \mathcal{H}$ and every $j = 1, 2, 3$. Substituting (3.1) into (3.5), we get

$$\begin{aligned} &E_0\Big[l_\theta(\theta_0, \Lambda_0; X) \sum_{i=1}^m h(C_i) W_i(\theta_0, \Lambda_0; X)\Big] \\ &\quad = E_0\Big[\sum_{k=1}^m \mathbf{h}^*(C_k) W_k(\theta_0, \Lambda_0; X) \sum_{i=1}^m h(C_i) W_i(\theta_0, \Lambda_0; X)\Big]. \end{aligned} \tag{3.6}$$

Since each C_i has the same marginal distribution Q, (3.6) becomes

$$\begin{aligned} &\sum_{i=1}^m \int h(C_i) E_0\Big[l_\theta(\theta_0, \Lambda_0; X) W_i(\theta_0, \Lambda_0; X) | C_i\Big] dQ(C_i) \\ &\quad = \sum_{i=1}^m \int h(C_i) E_0\Big[\sum_{k=1}^m \mathbf{h}^*(C_k) W_k(\theta_0, \Lambda_0; X) W_i(\theta_0, \Lambda_0; X) | C_i\Big] dQ(C_i). \end{aligned}$$

Thus (3.5) is equivalent to

$$\begin{aligned} &\int h(u) \sum_{i=1}^m E_0\Big[l_\theta(\theta_0, \Lambda_0; X) W_i(\theta_0, \Lambda_0; X) | C_i = u\Big] dQ(u) \\ &\quad = \int h(u) \sum_{i=1}^m E_0\Big[\sum_{k=1}^m \mathbf{h}^*(C_k) W_k(\theta_0, \Lambda_0; X) W_i(\theta_0, \Lambda_0; X) | C_i = u\Big] dQ(u). \end{aligned} \tag{3.7}$$

In view of the definition of $\mathbf{h}^*$ in (3.3), we know (3.7) is satisfied. This completes the proof.

Lemma 3.2. *The linear operator* A *defined by* (3.2) *is onto and continuously invertible.*

Proof. It suffices to show that A is injective and is the sum of a compact operator and a continuously invertible and surjective operator (see, for example, Theorem 4.25 in Rudin (1973), or Lemma 25.93 in van der Vaart (1998)).

We first consider the injectivity of A. If $Ah = 0$ for some $h \in \mathcal{H}$, then $\int h(Ah)dQ = 0$. Combining this with (3.1) and (3.2), we know $P_0(l_\Lambda(\theta_0, \Lambda_0)[h])^2 = 0$, and hence $l_\Lambda(\theta_0, \Lambda_0)[h] = 0$ a.s. $[P_0]$.

Considering $\Delta_i = 0$, Z_i near z_i, and C_i near c_1 from the right for $i = 1, \ldots, m$ in $l_\Lambda(\theta_0, \Lambda_0)[h] = 0$, and use the argument for deriving (2.3) to get

$$h(c_1)\Big[\sum_{i=1}^{m}\int p(\eta;\theta_{10},\theta_{20})e^{\theta_{30}z_i}(\eta_0+\eta_i)\prod_{j=1}^{m}e^{-e^{\theta_{30}z_j}(\eta_0+\eta_j)\Lambda_0(c_1)}d\eta\Big] = 0$$

for almost every $(c_1, z_1, \ldots, z_m)$ in the support of the distribution $(C_1, Z_1, \ldots, Z_m)$. Since the term in the square brackets of the preceding equation is positive for almost every $(c_1, z_1, \ldots, z_m)$, we know $h = 0$ *a.e.* $[Q]$.

We now show that A is the sum of a compact operator and a continuously invertible and surjective operator. In view of

$$\begin{aligned}(Ah)(u) &= h(u)\sum_{i=1}^{m}E_0(W_i^2(\theta_0,\Lambda_0;X)|C_i = u)\\ &\quad+\sum_{i=1}^{m}\sum_{\substack{j=1\\ j\neq i}}^{m}E_0(h(C_j)W_j(\theta_0,\Lambda_0;X)W_i(\theta_0,\Lambda_0;X)|C_i = u),\end{aligned}$$

we define $A_0 : \mathcal{H} \to \mathcal{H}$ by

$$(A_0h)(u) = h(u)\sum_{i=1}^{m}E_0(W_i^2(\theta_0,\Lambda_0;X)|C_i = u).$$

Since $\sum_{i=1}^{m} E_0(W_i^2(\theta_0, \Lambda_0; X)|C_i = u) > 0$ with probability 1, we know A_0 is onto and continuously invertible. Therefore, it suffices to show that $A - A_0$ is a compact operator. Note that $A - A_0$ is a linear operator with

$$\begin{aligned}&\|Ah - A_0h\|_{BV}\\ &= \Big\|\sum_{i=1}^{m}\sum_{\substack{j=1\\ j\neq i}}^{m}\int_{\tau_1}^{\tau_2}h(C_j)E_0(W_j(\theta_0,\Lambda_0;X)W_i(\theta_0,\Lambda_0;X)|C_i = u, C_j)dQ(C_j)\Big\|_{BV}\\ &\leq b\|h\|_{BV},\end{aligned}\tag{3.8}$$

for every $h \in \mathcal{H}$ and some constant $b > 0$. Using (3.8), Helly's Selection Lemma, and the Dominated Convergence Theorem, we know every sequence $(Ah_n - A_0h_n)_{n\geq 1}$ has a convergent subsequence if h_n in $\mathcal{H}$ satisfies $\|h_n\|_{BV} \leq 1$. This completes the proof.

Theorem 3.3. *I_0 is positive definite.*

Proof. Let $\nu = (\nu_1, \nu_2, \nu_3) \in \mathcal{R}^3$. Since $\nu I_0 \nu^T = P_0(\tilde{l}_0\nu^T)^2 \geq 0$, it suffices to show that $\nu I_0\nu^T = 0$ implies $\nu = 0$.

Suppose $\nu I_0 \nu^T = 0$, then $\tilde{l}_0 \nu^T = 0$ $a.s.$ $[P_0]$. Let $\delta_i \in \{0,1\}$, z_i be the point in the support of the distribution of Z_i, and c_i^* be the point given in **(A3)**. Considering $\Delta_i = \delta_i$, Z_i near z_i, and C_i near c_i^* from the right for $i = 1, \ldots, m$ in $\tilde{l}_0 \nu^T = 0$, and use the argument for deriving (2.3) to get

$$\left(\sum_{j=1}^{3} \nu_j \frac{\partial}{\partial \theta_j} - \sum_{i=1}^{m} (\sum_{j=1}^{3} \nu_j h_j^*(c_i^*)) \frac{\partial}{\partial y_i}\right)\Bigg|_{(\theta, y_1, \ldots, y_m) = (\theta_0, \Lambda_0(c_1^*), \ldots, \Lambda_0(c_m^*))}$$
$$\log \int p(\eta; \theta_1, \theta_2) \prod_{k=1}^{m} \left[1 - e^{-e^{\theta_3 z_k}(\eta_0 + \eta_k) y_k}\right]^{\delta_k} \left[e^{-e^{\theta_3 z_k}(\eta_0 + \eta_k) y_k}\right]^{1-\delta_k} d\eta = 0$$

for $(\delta_1, \ldots, \delta_m) \in \{0,1\}^m$ and almost every $(z_1, \ldots, z_m)$ in the support of the distribution of $(Z_1, \ldots, Z_m)$. Using **(A3)**, we know $\nu_1 = \nu_2 = \nu_3 = 0$. This completes the proof.

4. Rate of Convergence

With the consistency established in Subsection 2.3, we now apply empirical process theory to study the rate of convergence for $(\hat{\theta}_K, \hat{\Lambda}_K)$ under the assumption that $(\hat{\theta}_K, \hat{\Lambda}_K) \in \mathcal{N}_{\theta_0} \times \mathcal{L}_0$, where $\mathcal{N}_{\theta_0}$ is a neighborhood of θ_0 and $\mathcal{L}_0 = \{\Lambda \in \mathcal{L} \mid 1/M \leq \Lambda(\tau_1) \leq \Lambda(\tau_2) \leq M\}$. Define the profile likelihood for θ,

$$pL_K(\theta) = \sup_{\Lambda \in \mathcal{L}_0} L_K(\theta, \Lambda),$$

where $L_K(\theta, \Lambda)$ is the full likelihood given by (2.4). For every fixed θ, denote by $\hat{\Lambda}_\theta$ a random element at which the supremum in the definition of pL_K is achieved. The existence of $\hat{\Lambda}_\theta$ can be established by the argument in the proof of Theorem 2.2.

The main result of this section is the following.

Theorem 4.1. *For every random sequence* $\tilde{\theta}_K \xrightarrow{P} \theta_0$,

$$\|\hat{\Lambda}_{\tilde{\theta}_K} - \Lambda_0\|_{2,Q} = O_P(\|\tilde{\theta}_K - \theta_0\| + K^{-\frac{1}{3}}), \tag{4.1}$$

where $\|\hat{\Lambda}_{\tilde{\theta}_K} - \Lambda_0\|_{2,Q} = (\int (\hat{\Lambda}_{\tilde{\theta}_K} - \Lambda_0)^2 dQ)^{1/2}$, *and* Q *is the marginal distribution of the examination variable* C_i.

The proof of Theorem 4.1 is at the end of this section, when a series of lemmas is established. Lemma 4.2 provides an upper bound for the entropy-with-bracketing integral for the class of log-likelihood functions. Lemmas 4.4 and 4.5 concern the modulus of continuity of the empirical process indexed by the log-likelihoods.

Let $\Psi = \{\log lik(\theta,\Lambda) \mid (\theta,\Lambda) \in \mathcal{N}_{\theta_0} \times \mathcal{L}_0\}$. For any probability measure P on the sample space Ω, let $L^2(P) = \{g \mid Pg^2 < \infty\}$ and $\|g\|_{2,P} = (Pg^2)^{1/2}$ for $g \in L^2(P)$. Given any subclass $\mathcal{C}$ of $L^2(P)$, we define the bracketing number

$$N_{[\,]}(\varepsilon,\mathcal{C},L_2(P)) = \min\{N \mid \text{there exists } f_1^L, f_1^U, \ldots, f_N^L, f_N^U \text{ such that}$$
$$\|f_n^L - f_n^U\|_{2,P} < \varepsilon, \text{ and for each } f \in \mathcal{C}, f_n^L \le f \le f_n^U \text{ for some } 1 \le n \le N\},$$

and the bracketing integral

$$J_{[\,]}(\delta,\mathcal{C},L_2(P)) = \int_0^\delta \sqrt{1+\log N_{[\,]}(\varepsilon,\mathcal{C},L_2(P))}d\varepsilon.$$

Lemma 4.2. $\log N_{[\,]}(\varepsilon,\Psi,L_2(P_0)) = O(1/\varepsilon)$ *as* ε *decreases to* 0.

Proof. Because of the monotonicity of the elements in $\mathcal{L}_0$, we know $N_{[\,]}(\varepsilon,\mathcal{L}_0, L_2(Q)) \le e^{c/\varepsilon}$ for some constant $c > 0$. (See, for example, Theorem 2.7.5 of van der Vaart and Wellner (1996)). This implies that there exists a sequence of functions $\{\tilde\Lambda_j^L, \tilde\Lambda_j^U, j = 1,\ldots,J\}$, where $J = O(e^{c/\varepsilon})$, such that $\|\tilde\Lambda_j^U - \tilde\Lambda_j^L\|_{2,Q} < \varepsilon$, and for each $\Lambda \in \mathcal{L}_0$, $\tilde\Lambda_j^L \le \Lambda \le \tilde\Lambda_j^U$ for some $1 \le j \le J$. Let $\Lambda_j^L = \tilde\Lambda_j^L - \varepsilon$ and $\Lambda_j^U = \tilde\Lambda_j^U + \varepsilon$. Then, $\|\Lambda_j^U - \Lambda_j^L\|_{2,Q} < 3\varepsilon$. Since elements of $\mathcal{L}_0$ are uniformly bounded away from zero, we can choose ε small enough that each Λ_j^L also stays away from zero.

For each $\Lambda \in \mathcal{L}_0$, we assign one pair Λ_j^L and Λ_j^U so that $\tilde\Lambda_j^L \le \Lambda \le \tilde\Lambda_j^U$. For $(\theta,\Lambda) \in \mathcal{N}_{\theta_0} \times \mathcal{L}_0$, we define

$$l_{j,\theta}^L(x) = \log \int p(\eta;\theta_1,\theta_2) \prod_{i=1}^m \Big[1 - e^{-e^{\theta_3 z_i}(\eta_0+\eta_i)\Lambda_j^L(c_i)}\Big]^{\delta_i} \Big[e^{-e^{\theta_3 z_i}(\eta_0+\eta_i)\Lambda_j^U(c_i)}\Big]^{1-\delta_i} d\eta,$$
$$l_{j,\theta}^U(x) = \log \int p(\eta;\theta_1,\theta_2) \prod_{i=1}^m \Big[1 - e^{-e^{\theta_3 z_i}(\eta_0+\eta_i)\Lambda_j^U(c_i)}\Big]^{\delta_i} \Big[e^{-e^{\theta_3 z_i}(\eta_0+\eta_i)\Lambda_j^L(c_i)}\Big]^{1-\delta_i} d\eta.$$

Here $x = (c_1,\delta_1,z_1,\ldots,c_m,\delta_m,z_m)$.

Consider the function

$$f(\alpha;x) = \log \int p(\eta;\theta_1,\theta_2) \prod_{i=1}^m \Big[1 - e^{-e^{\theta_3 z_i}(\eta_0+\eta_i)y_i}\Big]^{\delta_i} \Big[e^{-e^{\theta_3 z_i}(\eta_0+\eta_i)y_i}\Big]^{1-\delta_i} d\eta, \tag{4.2}$$

where $\alpha = (\theta, y_1,\ldots,y_m)$ is in $\mathcal{N}_{\theta_0} \times [1/M, M]^m$, and x in the range of X.

Let $\alpha' = (\theta', (\delta_1\Lambda_j^U + (1-\delta_1)\Lambda_j^L)(c_1), \ldots, (\delta_m\Lambda_j^U + (1-\delta_m)\Lambda_j^L)(c_m))$ and $\alpha'' = (\theta, \Lambda(c_1),$
$\ldots, \Lambda(c_m))$. Applying the Mean Value Theorem, there exists an intermediate point $\tilde\alpha$ between α' and α'' such that

$$l_{j,\theta'}^U(x) - \log lik(\theta,\Lambda;x)$$

$$= f(\alpha^{'};x) - f(\alpha^{''};x)$$
$$= \sum_{i=1}^{3} f_i(\tilde{\alpha})(\theta_i^{'}-\theta_i)+\sum_{i=1}^{m} f_{i+3}(\tilde{\alpha})\Big[\delta_i(\Lambda_j^U-\Lambda)(c_i)+(1-\delta_i)(\Lambda_j^L-\Lambda)(c_i)\Big]. \quad (4.3)$$

Here f_i denotes the partial derivative of f with respect to α_i with $(\alpha_1,\alpha_2,\alpha_3) = \theta$, and $(\alpha_4,\ldots,\alpha_{m+1})=(y_1,\ldots,y_m)$. We note that, for $4\le i\le m+3$, $f_i(\alpha;x)/(2\delta_i-1)$ is positive and uniformly bounded in $(\alpha;x)$. It follows from (4.3) that

$$l_{j,\theta'}^U(x) - \log lik(\theta,\Lambda;x)$$
$$\geq -b_0\|\theta^{'}-\theta\| + b_1\sum_{i=1}^{m}[2\delta_i-1]\Big[\delta_i(\Lambda_j^U-\Lambda)(c_i)+(1-\delta_i)(\Lambda_j^L-\Lambda)(c_i)\Big] \quad (4.4)$$

for some constants $b_0, b_1 > 0$.

By the definition of $(\Lambda_j^L,\Lambda_j^U)$, we know form (4.4) that $l_{j,\theta'}^U(x)-\log lik(\theta,\Lambda;x) \geq -b_0\|\theta^{'}-\theta\|+b_1\varepsilon$. Similarly, we have $l_{j,\theta'}^L(x)-\log lik(\theta,\Lambda;x) \le b_0\|\theta^{'}-\theta\|-b_2\varepsilon$ for some constant $b_2 > 0$.

Let $\theta^{(1)},\ldots,\theta^{(N)}$ be points in $\mathcal{N}_{\theta_0}$ such that for every $\theta\in\mathcal{N}_{\theta_0}$, $\|\theta-\theta^{(n)}\| \le \min\{b_1\varepsilon/b_0, b_2\varepsilon/b_0\}$ for some $1\le n\le N$. Therefore, for every (θ,Λ) in $\mathcal{N}_{\theta_0}\times\mathcal{L}_0$, there exist $\theta^{(n)}$, Λ_j^L and Λ_j^U such that

$$l_{j,\theta^{(n)}}^L \le \log lik(\theta,\Lambda) \le l_{j,\theta^{(n)}}^U. \quad (4.5)$$

Because $\mathcal{N}_{\theta_0}\subset\mathcal{R}^3$, we note that N can be on the order of $O(1/\varepsilon^3)$.

Furthermore, we know from the Mean Value Theorem that

$$\|l_{j,\theta^{(n)}}^U - l_{j,\theta^{(n)}}^L\|_{2,P}^2 \le E\Big(b_3\sum_{i=1}^{m}|\Lambda_j^U(c_i)-\Lambda_j^L(c_i)|^2\Big)$$
$$= b_3 m\|\Lambda_j^U-\Lambda_j^L\|_{2,Q}^2 < b_3 m(3\varepsilon)^2 \quad (4.6)$$

for some $b_3>0$. It follows from (4.5) and (4.6) that $N_{[\,]}(\varepsilon,\Psi,L_2(P))$ is of the order $NJ = O(e^{c/\varepsilon}/\varepsilon^3)$, and hence $\log N_{[\,]}(\varepsilon,\Psi,L_2(P)) = O(1/\varepsilon)$. This completes the proof.

Let P_1 and P_2 denote the distribution of $(C_1,\ldots,C_m)$ and $(C_1,\ldots,C_m,Z_1,\ldots,Z_m)$ respectively. We consider the family of conditional log-densities of X given $(C_1,\ldots,C_m)=(c_1,\ldots,c_m)$,

$$g(\alpha;x) = \log\Big[lik(\theta,\Lambda;x)\frac{dP_2(c_1,\ldots,c_m,z_1,\ldots,z_m)}{dP_1(c_1,\ldots,c_m)}\Big], \quad (4.7)$$

parameterized by $\alpha=(\theta,\Lambda(c_1),\ldots,\Lambda(c_m))\in\mathcal{R}^{m+3}$. Here $x=(c_1,\delta_1,z_1,\ldots,c_m,\delta_m,z_m)$.

Denote the first and the second derivative of g relative to α by $\dot{g}$ and $\ddot{g}$ respectively, and let $\Sigma_{(c_1,\ldots,c_m)} \equiv E_0(\ddot{g}(\alpha_0; X)|C_1 = c_1, \ldots, C_m = c_m)$. Here $\alpha_0 = (\theta_0, \Lambda_0(c_1), \ldots, \Lambda_0(c_m))$. Since (4.7) defines a parametric model, we have

$$-u\Sigma_{(c_1,\ldots,c_m)}u^T = E_0\bigg((\dot{g}(\alpha_0; X)u^T)^2 \mid C_1 = c_1, \ldots, C_m = c_m\bigg) \geq 0,$$

for every $u = (u_1, \ldots, u_{m+3}) \in \mathcal{R}^{m+3}$.

Lemma 4.3. $\Sigma_{(c_1^*,\ldots,c_m^*)}$ *is negative definite.*

Proof. It suffices to show that $u\Sigma_{(c_1^*,\ldots,c_m^*)}u^T = 0$ implies $u = 0$. Let $\alpha_0^* = (\theta_0, \Lambda_0(c_1^*), \ldots, \Lambda_0(c_m^*))$. Suppose $u\Sigma_{(c_1^*,\ldots,c_m^*)}u^T = 0$, then

$$\dot{g}(\alpha_0^*; c_1^*, \delta_1, z_1, \ldots, c_m^*, \delta_m, z_m)u^T = 0 \tag{4.8}$$

for almost every $\delta_i \in \{0, 1\}$ and z_i in the support of the distribution of Z_i. Noting that (4.8) is (2.1) precisely, we get $u = 0$ by **(A3)**. This completes the proof.

Lemma 4.4. *There exists a constant* $b > 0$ *such that*

$$P_0\bigg(\log lik(\theta, \Lambda) - \log lik(\theta, \Lambda_0)\bigg) \leq b(-\|\Lambda - \Lambda_0\|_{2,Q}^2 + \|\theta - \theta_0\|^2)$$

for every $(\theta, \Lambda) \in \mathcal{N}_{\theta_0} \times \mathcal{L}_0$.

Proof. It suffices to show that

$$P_0\bigg(\log lik(\theta_0, \Lambda_0) - \log lik(\theta, \Lambda_0)\bigg) \leq b\|\theta - \theta_0\|^2, \tag{4.9}$$

$$P_0\bigg(\log lik(\theta, \Lambda) - \log lik(\theta_0, \Lambda_0)\bigg) \leq -b\|\Lambda - \Lambda_0\|_{2,Q}^2. \tag{4.10}$$

A Taylor series argument in θ can be used to verify (4.9). We prove (4.10).

Using the Taylor's expansion of g around α_0, we know

$$\begin{aligned}
&P_0\bigg(\log lik(\theta, \Lambda) - \log lik(\theta_0, \Lambda_0)\bigg) \\
&\quad = E_0 E_0\bigg(g(\alpha; X) - g(\alpha_0; X) \mid C_1, \ldots, C_m\bigg) \\
&\quad = E_0\bigg(E_0(\dot{g}(\alpha_0; X)|C_1, \ldots, C_m)(\alpha - \alpha_0)^T \\
&\qquad + (\alpha - \alpha_0)\Sigma_{(C_1,\ldots,C_m)}(\alpha - \alpha_0)^T + o(\|\alpha - \alpha_0\|^2)\bigg) \\
&\quad = E_0\bigg(l_\theta(\theta_0, \Lambda_0)(X)(\theta - \theta_0)^T + l_\Lambda(\theta_0, \Lambda_0)[\Lambda - \Lambda_0](X)\bigg)
\end{aligned}$$

$$+E_0\Big((\alpha-\alpha_0)\Sigma_{(C_1,\ldots,C_m)}(\alpha-\alpha_0)^T+o(\|\alpha-\alpha_0\|^2)\Big). \tag{4.11}$$

Recalling that l_θ and l_Λ are the score functions and using Rayleigh's principle (see, for example, Theorem 6.37 in Stephen, Arnold and Lawrence (1997)), we know from (4.11) that

$$P_0\Big(\log lik(\theta,\Lambda)-\log lik(\theta_0,\Lambda_0)\Big)\le E_0\Big(\lambda(C_1,\ldots,C_m)\|\alpha-\alpha_0\|^2+o(\|\alpha-\alpha_0\|^2)\Big), \tag{4.12}$$

where $\lambda(C_1,\ldots,C_m)$ is the largest eigenvalue of $\Sigma_{(C_1,\ldots,C_m)}$. Noting that λ is continuous at $(c_1^*,\ldots,c_m^*)$ and using Lemma 4.3, we know λ has a negative upper bound on some neighborhood of $(c_1^*,\ldots,c_m^*)$. Combining this with the negative semi-definiteness of $\Sigma_{(C_1,\ldots,C_m)}$ and (4.12), we obtain (4.10). This completes the proof.

Lemma 4.5. *Let $\phi_K(\delta)=\sqrt{\delta}(1+\sqrt{\delta}/(\delta^2\sqrt{K}))$ and let E^* denote outer expectation. Then there exists a constant $B>0$ such that*

$$E^*\sup_{\theta\in\mathcal{N}_{\theta_0},\ \Lambda\in\mathcal{L}_0,\ \|\theta-\theta_0\|<\delta,\ \|\Lambda-\Lambda_0\|_{2,Q}<\delta}|G_K(\log lik(\theta,\Lambda)-\log lik(\theta_0,\Lambda_0))|\le B\phi_K(\delta),$$

for δ sufficiently small.

Proof. We first note that all elements of Ψ are uniformly bounded. Using the Mean Value Theorem, there exists a constant $b>0$ such that

$$P_0\big(\log lik(\theta,\Lambda)-\log lik(\theta,\Lambda_0)\big)^2\le b(\|\Lambda-\Lambda_0\|_{2,Q}^2+\|\theta-\theta_0\|^2) \tag{4.13}$$

for every $(\theta,\Lambda)\in\mathcal{N}_{\theta_0}\times\mathcal{L}_0$.

Furthermore, we know from Lemma 4.2 that

$$J_{[\,]}(\delta,\Psi,L_2(P))=O(\sqrt{\delta}) \tag{4.14}$$

as δ decreases to 0. It follows from (4.13), (4.14), and Lemma 3.3 of Murphy and van der Vaart (1999) that the proof is complete.

Proof of Theorem 4.1. In view of Lemma 4.4 and Lemma 4.5, we know the conditions of Theorem 3.2 of Murphy and van der Vaart (1999) are satisfied for $\phi_K(\delta)=\sqrt{\delta}(1+\sqrt{\delta}/(\delta^2\sqrt{K}))$. Since $K^{2/3}\phi_K(K^{-1/3})=2\sqrt{K}$, we know $\|\hat{\Lambda}_{\tilde{\theta}_K}-\Lambda_0\|_{2,Q}=O_P(\|\tilde{\theta}_K-\theta_0\|+K^{-1/3})$ for every random sequence $\tilde{\theta}_K\xrightarrow{P}\theta_0$. This completes the proof.

5. Profile Likelihood Theory

In this section, we focus our attention on the estimation of θ and present a profile likelihood theory.

Formally, the efficient score function $\tilde{l}_0$ is the derivative at $\nu = \theta_0$ of the log-likelihood function evaluated at the path $\nu \mapsto (\nu, \Lambda_0 + \mathbf{h}^*(\theta_0 - \nu)^T)$. The so-called least favorable submodel refers to this path. However, the second coordinate of the preceding path may not lie in the space $\mathcal{L}_0$ defined in Section 4. We now modify and replace this path to obtain an approximately least-favorable submodel.

Let $\phi : [0, M] \to [0, \infty)$ be defined by

$$\phi(y) = \begin{cases} 0 & \text{for } 0 \le y < \frac{1}{M}, \\ \frac{y - M^{-1}}{\Lambda_0(\tau_1) - M^{-1}} & \text{for } \frac{1}{M} \le y < \Lambda_0(\tau_1), \\ 1 & \text{for } \Lambda_0(\tau_1) \le y < \Lambda_0(\tau_2), \\ \frac{M-y}{M - \Lambda_0(\tau_2)} & \text{for } \Lambda_0(\tau_2) \le y \le M. \end{cases}$$

For fixed (θ, Λ) and $\nu = (\nu_1, \nu_2, \nu_3) \in \mathcal{R}^3$, we define

$$\mathbf{\Lambda}_\nu(\theta, \Lambda)(t) = \Lambda(t) + \phi(\Lambda(t))(\mathbf{h}^* \circ \Lambda_0^{-1})(\Lambda(t))(\theta - \nu)^T. \tag{5.1}$$

Recall that a real-valued function g is *Lipschitz* if there exists a number L such that $|g(u_1) - g(u_2)| \le L|u_1 - u_2|$ for every u_1 and u_2. The least such number L is denoted by $\|g\|_{Lip}$. Because of **(A4)**, the mapping

$$u \mapsto \sum_{i=1}^{m} E_0(l_\theta(\theta_0, \Lambda_0; X) W_i(\theta_0, \Lambda_0; X) | C_i = u)$$

is Lipschitz. Let $\| \sum_i E_0(l_\theta(\theta_0, \Lambda_0; X) W_i(\theta_0, \Lambda_0; X) | C_i = \cdot\ )\|_{Lip} = L_0$, and let $\mathcal{H}_0$ be the closed linear span of $\{h \in \mathcal{H} \mid h \text{ is Lipschitz with } \|h\|_{Lip} \le L_0\}$. We note that the proofs of Theorem 3.1 and Lemma 3.2 indicate that the operator $A : \mathcal{H}_0 \to \mathcal{H}_0$ is onto and continuously invertible. This shows that the functions $\{h_j^*; j = 1, 2, 3\}$ are bounded and Lipschitz. Based on this and **(A4)**, the following result can be obtained straightforwardly.

Lemma 5.1. $\mathbf{\Lambda}_\nu(\theta, \Lambda)(\cdot)$ *given by* (5.1) *defines a cumulative hazard function in* $\mathcal{L}_0$ *for every* ν *sufficiently close to* θ.

Using Lemma 5.1, we introduce the approximately least-favorable submodel specified by the log-likelihood $\nu \mapsto l(\nu, \theta, \Lambda; X) \equiv \log lik(\nu, \mathbf{\Lambda}_\nu(\theta, \Lambda); X)$. We denote respectively the first and the second derivative of l relative to ν by $\dot{l}(\nu, \theta, \Lambda; X)$ and $\ddot{l}(\nu, \theta, \Lambda; X)$. Noting that

$$\dot{l}(\nu, \theta, \Lambda; X) = l_\theta(\nu, \mathbf{\Lambda}_\nu(\theta, \Lambda))(X) - l_\Lambda(\nu, \mathbf{\Lambda}_\nu(\theta, \Lambda))[(\phi \mathbf{h}^* \circ \Lambda_0^{-1})(\Lambda)](X), \tag{5.2}$$

we know

$$\dot{l}(\theta_0, \theta_0, \Lambda_0; X) = \tilde{l}_0(X), \tag{5.3}$$

the efficient score function for θ at (θ_0, Λ_0). Furthermore, noting that $\nu \mapsto l(\nu, \theta_0, \Lambda_0)$ is a smooth parametric submodel, its derivatives at θ_0 satisfies

$$P_0 \ddot{l}(\theta_0, \theta_0, \Lambda_0) = -P_0 \dot{l}^T(\theta_0, \theta_0, \Lambda_0)\dot{l}(\theta_0, \theta_0, \Lambda_0) = -I_0. \tag{5.4}$$

Lemma 5.2. *The class of functions* $\{\dot{l}(\nu, \theta, \Lambda) \mid (\nu, \theta, \Lambda) \in \mathcal{N}_{\theta_0} \times \mathcal{N}_{\theta_0} \times \mathcal{L}_0\}$ *is a uniformly bounded Donsker class, and the class of functions* $\{\ddot{l}(\nu, \theta, \Lambda) \mid (\nu, \theta, \Lambda) \in \mathcal{N}_{\theta_0} \times \mathcal{N}_{\theta_0} \times \mathcal{L}_0\}$ *is a uniformly bounded Glivenko-Cantelli class.*

Readers are referred to van der Vaart and Wellner (1996) for the definitions of a Donsker class and a Glivenko-Cantelli class. The proof for Lemma 5.2 is technical and hence omitted. Readers can find it in Chang, Hsiung and Wen (2002).

Lemma 5.3. *For every random sequence* $\tilde{\theta}_K \xrightarrow{P} \theta_0$,

$$P_0 \dot{l}(\theta_0, \tilde{\theta}_K, \hat{\Lambda}_{\tilde{\theta}_K}) = o_P(\|\tilde{\theta}_K - \theta_0\| + K^{-\frac{1}{2}}). \tag{5.5}$$

Proof. Since $\dot{l}(\theta, \theta, \Lambda)$ is a score function for the model indexed by (θ, Λ), we have $P_{\theta,\Lambda}\dot{l}(\theta, \theta, \Lambda) = 0$ for every (θ, Λ). Differentiating this identity relative to θ yields

$$P_{\theta,\Lambda} l_\theta^T(\theta, \Lambda)\dot{l}(\theta, \theta, \Lambda) + P_{\theta,\Lambda}\ddot{l}(\theta, \theta, \Lambda) + \frac{\partial}{\partial \upsilon}\Big|_{\upsilon=\theta} P_{\theta,\Lambda}\dot{l}(\theta, \upsilon, \Lambda) = 0,$$

where $l_\theta(\theta, \Lambda)$ is the score function for θ. Evaluating this at $(\theta, \Lambda) = (\theta_0, \Lambda_0)$ gives

$$-\frac{\partial}{\partial \upsilon}\Big|_{\upsilon=\theta_0} P_0 \dot{l}(\theta_0, \upsilon, \Lambda_0) = P_0 l_\theta^T(\theta_0, \Lambda_0)\dot{l}(\theta_0, \theta_0, \Lambda_0) + P_0 \ddot{l}(\theta_0, \theta_0, \Lambda_0) = I_0 - I_0 = 0.$$

Here (5.4) and the fact that $\dot{l}(\theta_0, \theta_0, \Lambda_0)(= \tilde{l}_0)$ is orthogonal to every Λ-score are used to get the second to last equality. Thus,

$$P_0(\dot{l}(\theta_0, \theta, \Lambda) - \dot{l}(\theta_0, \theta_0, \Lambda)) = P_0\left(\frac{\partial}{\partial \upsilon}\Big|_{\upsilon=\theta_*} \dot{l}(\theta_0, \upsilon, \Lambda) - \frac{\partial}{\partial \upsilon}\Big|_{\upsilon=\theta_0} \dot{l}(\theta_0, \upsilon, \Lambda_0)\right)(\theta - \theta_0)^T$$

for an intermediate point θ_* between θ and θ_0. Letting $\theta = \tilde{\theta}_K$ and $\Lambda = \hat{\Lambda}_{\tilde{\theta}_K}$, and using (4.1) and the Mean Value Theorem, we know it suffices to verify

$$P_0 \dot{l}(\theta_0, \theta_0, \hat{\Lambda}_{\tilde{\theta}_K}) = o_P(\|\tilde{\theta}_K - \theta_0\| + K^{-\frac{1}{2}}) \tag{5.6}$$

to establish (5.5).

Noting that $P_{\theta_0,\Lambda}\dot{l}(\theta_0, \theta_0, \Lambda) = 0$, we have

$$P_0 \dot{l}(\theta_0, \theta_0, \Lambda) = (P_0 - P_{\theta_0,\Lambda})(\dot{l}(\theta_0, \theta_0, \Lambda) - \dot{l}(\theta_0, \theta_0, \Lambda_0)) + (P_0 - P_{\theta_0,\Lambda})\dot{l}(\theta_0, \theta_0, \Lambda_0). \tag{5.7}$$

We now explain, without giving the details, that both terms on the right-hand side are bounded by a multiple of $\|\Lambda - \Lambda_0\|_{2,Q}^2$. The desired bound for the first term is obtained by means of the Mean Value Theorem and the Cauchy-Schwarz inequality. The bound for the second term is obtained by means of the second-order Taylor expansion for $(\Lambda(C_1), \ldots, \Lambda(C_m)) \mapsto lik(\theta_0, \Lambda; X)$ around $(\Lambda_0(C_1), \ldots, \Lambda_0(C_m))$, and the fact $\dot{l}(\theta_0, \theta_0, \Lambda_0)$ is the efficient score $\tilde{l}_0$.

Applying the rate of convergence on $\hat{\Lambda}_{\tilde{\theta}_K}$ given by Theorem 4.1 to (5.7), we obtain $P_0 \dot{l}(\theta_0, \theta_0, \hat{\Lambda}_{\tilde{\theta}_K}) = O_P(\|\tilde{\theta}_K - \theta_0\|^2 + K^{-2/3})$. This is more than required and thus the proof is complete.

Theorem 5.4. *For every random sequence* $\tilde{\theta}_K \xrightarrow{P} \theta_0$,

$$\log pL_K(\tilde{\theta}_K) - \log pL_K(\theta_0) = (\tilde{\theta}_K - \theta_0) \sum_{k=1}^{K} \tilde{l}_0^T(X_k) - \frac{1}{2} K(\tilde{\theta}_K - \theta_0) I_0 (\tilde{\theta}_K - \theta_0)^T$$
$$+ o_{P_0}(\sqrt{K}\|\tilde{\theta}_K - \theta_0)\| + 1)^2. \tag{5.8}$$

Proof. We apply Theorem 1 of Murphy and van der Vaart (2000) to prove this theorem. It is easy to see that the functions $(\nu, \theta, \Lambda) \mapsto \dot{l}(\nu, \theta, \Lambda; X)$ and $(\nu, \theta, \Lambda) \mapsto \ddot{l}(\nu, \theta, \Lambda; X)$ are continuous at $(\theta_0, \theta_0, \Lambda_0)$ for P_0−almost every X. For every random sequence $\tilde{\theta}_K \xrightarrow{P} \theta_0$, (4.1) implies that $\hat{\Lambda}_{\tilde{\theta}_K} \xrightarrow{P} \Lambda_0$. In view of (5.1), (5.3), (5.5) and Lemma 5.2, we know all conditions of Theorem 1 of Murphy and van der Vaart (2000) are satisfied. Thus the proof is complete.

Using the consistency of $\hat{\theta}_K$, the invertibility of the efficient Fisher information matrix I_0, and the second order expansion of the profile likelihood (5.8), we obtain the following three theorems immediately from the profile likelihood theory of Murphy and van der Vaart (2000).

Theorem 5.5. *The NPMLE* $\hat{\theta}_K$ *is asymptotically normal and asymptotically efficient at* (θ_0, Λ_0); *that is,*

$$\sqrt{K}(\hat{\theta}_K - \theta_0) = I_0^{-1} \sqrt{K} P_K \tilde{l}_0^T + o_{P_0}(1) \xrightarrow{d} N(0, I_0^{-1}).$$

Theorem 5.6. *Under the null hypothesis* $\mathbf{H}_0 : \theta = \theta_0$, *the profile likelihood ratio statistic*

$$lrt_K(\theta_0) \equiv 2 \log \frac{pL_K(\hat{\theta}_K)}{pL_K(\theta_0)},$$

is asymptotically chi-squared with three degrees of freedom. The region $\{\theta \mid lrt_K(\theta) \leq \chi^2_{3,1-\alpha}\}$ *is an associated confidence region of asymptotic level* $1 - \alpha$.

Theorem 5.7. *For all sequences* $\nu_K \xrightarrow{P} \nu \in \mathcal{R}^3$ *and* $h_K \xrightarrow{P} 0$ *such that* $(\sqrt{K}h_K)^{-1} = O_P(1)$,

$$-2\frac{\log pL_K(\hat{\theta}_K + h_K\nu_K) - \log pL_K(\hat{\theta}_K)}{Kh_K^2} \xrightarrow{P} \nu I_0 \nu^T.$$

6. A Simulation Study

This section reports a simulation study that illustrates the numerical performance of the profile likelihood ratio statistic. The main task is to find $\sup_{\Lambda\in\mathcal{L}_0} L_K(\theta_0, \Lambda)$ and $\sup_{\theta\in\Theta_0,\Lambda\in\mathcal{L}_0} L_K(\theta, \Lambda)$. This is an optimization problem with the objective functions defined on a set of high dimension. In order to alleviate the computation burden, we consider sieve estimates. Let $b_1 < \cdots < b_N$ in $[\tau_1, \tau_2]$. We consider the estimates $\hat{\Lambda}$ that are in $\mathcal{L}_1 \subset \mathcal{L}$, which comprises step functions with possible jump points b_i. A function Λ in $\mathcal{L}_1$ can thus be identified with a nonnegative vector $v = (v_1, \ldots, v_N)$ with $v_j = \Lambda(b_j) - \Lambda(b_{j-1})$ for $j = 1, \ldots, N$; in fact, $\Lambda(c) = \sum_{j:b_j\le c} v_j$. The sieve maximum likelihood estimates of θ_0 and Λ_0 is the $\hat{\theta}$ and $\hat{\Lambda}$, represented by $\hat{v}$, that maximizes

$$\Phi_1(\theta, v) = \sum_{k=1}^{K} \log \int p(\eta; \theta_1, \theta_2) \prod_{i=1}^{m} \left[1 - e^{-e^{\theta_3 Z_{ik}}(\eta_0+\eta_i)\sum_{j:b_j\le C_{ik}} v_j}\right]^{\Delta_{ik}} \times \left[e^{-e^{\theta_3 Z_{ik}}(\eta_0+\eta_i)\sum_{j:b_j\le C_{ik}} v_j}\right]^{1-\Delta_{ik}} d\eta$$

subject to $\theta \in \Theta$ and $v \in (0, \infty)^N$.

Let $\xi = (\xi_1, \ldots, \xi_{N+3}) = (\log\theta_1, \log\theta_2, \theta_3, \log v_1, \ldots, \log v_N)$. Consider the bijective transform $\Phi(\xi) = \Phi_1(\theta, v)$ and denote the gradient of Φ relative to ξ by Φ'; namely, $\Phi' = (\frac{\partial\Phi}{\partial\xi_1}, \ldots, \frac{\partial\Phi}{\partial\xi_{N+3}})$. The following algorithm based on gradient method is used to find the estimates.

(1) Choose a starting point $\xi^{(1)}$.
(2) Set $J = 1$.
(3) Set $n = 1$.
(4) Let $\tilde{\xi} = \xi^{(J)} + 2^{-n}\Phi'(\xi^{(J)})$.
(5) If $\Phi(\xi^{(J)}) > \Phi(\tilde{\xi})$, then set $n = n + 1$ and go back to (4).
(6) If $\Phi(\xi^{(J)}) \le \Phi(\tilde{\xi})$, then set $\xi^{(J+1)} = \tilde{\xi}$.
(7) $J = J + 1$.
(8) Repeat (3)$\sim$(7) for a suitable number M of iterations once the evidence of convergence occurs.
(9) Set $(\hat{\theta}, \hat{v}) = (e^{\xi_1^{(M)}}, e^{\xi_2^{(M)}}, \xi_3^{(M)}, e^{\xi_4^{(M)}}, \ldots, e^{\xi_{N+3}^{(M)}})$.

We generate data with $\theta_0 = (1, 1, 0.5)$, $\Lambda_0(t) = \log(100/(100-t))$, $Z_1 \sim unif\{0,1\}, C_1 \sim unif(1,99)$, $m = 3$, and $K = 400$. Our study consists of 100 replicates. The parameters for the sieve are $N = 98$ and $b_i = i$ for $i = 1, \ldots, N$. In applying the above algorithm, we take the starting value $\xi^{(1)} = (0, 0, 0.5, \log(1/80), \log(2/80), \ldots, \log(98/80))$ to conduct the likelihood ratio test. Our simulation study seems to suggest that the χ^2 approximation works well for the profile likelihood ratio statistic.

The following Table 1 reports the theoretical and simulated critical values (CV) and rejection rate (RR) under the null hypothesis $\mathbf{H}_0$, and Figure 1 is the Q-Q plot for $lrt_{400}(\theta_0)$ versus χ^2_3.

Table 1. Comparison of theoretical and simulated critical values (CV) and rejection rate (RR) under H_0.

Significance Level (%)	CV for χ_3	CV for $lrt_{900}(\theta_0)$	RR (%)
90	0.5844	0.7058	93
80	1.0052	1.1144	82
70	1.4237	1.5589	72
60	1.8692	1.9569	61
50	2.3660	2.5947	51
40	2.9462	3.3848	47
30	3.6649	3.9840	36
20	4.6416	4.8706	21
10	6.2514	6.6243	13
5	7.8147	7.8293	6
1	11.3449	9.1968	0

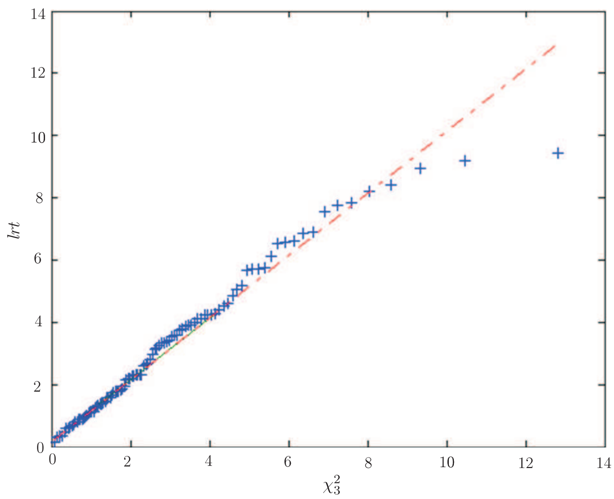

Figure 1. Q.Q plot for $lrt_{400}(\theta_0)$ v.s. χ^2_3.

We also studied this problem by a nonlinear conjugate gradient method with a strong global convergence property, as developed by Dai and Yuan (1999). The results from this method are omitted, because they are similar to those in Table 1.

7. Discussion

With the introduction of the identifiability assumption **(A3)**, we have established a profile likelihood theory for the correlated gamma-frailty model with current status family data. Specifically, we have obtained a quadratic expansion of the profile likelihood for the parameters of interest, and derived from it the asymptotic normality and efficiency of the NPMLE, the asymptotic distribution of the profile likelihood ratio statistic, and a consistent estimate of the covariance matrix of the NPMLE. This approach may also be useful in other models with family data.

In this paper, we allowed different examination times for different members in the same family. We note that our argument can be used to obtain a profile likelihood theory for the case that all the members in the same family share a common examination time. This situation arises, for example, when both eyes of a person are examined at the same time for a given disease. In fact, in this case, the efficient score has the closed form (3.4).

To make these methods useful in applications, we need to find good approximations to $\hat{\Lambda}, \hat{\theta}_1, \hat{\theta}_2$, and $\hat{\theta}_3$, and to know the numerical performance of the NPMLE. The simulation study in Section 6 represents our initial effort in this direction; although it is encouraging, a thorough study is needed and is underway. The main challenges come from the fact that the log-likelihood $\Phi_1(\theta, v)$ in Section 6 is not concave, its domain of definition has high dimensionality, and the likelihood itself is not concave. The latter can leave the likelihood ratio statistic based confidence region not convex (see Lemma A.1 of Murphy and van der Vaart (1997); in fact, our (unreported) simulation study does indicate this possibility. We note that, in the study of Cox's model with current status data, Huang (1996) made significant use of the concavity of the log-likelihood with respect to the cumulative hazard function. Faced with the above mentioned difficulties, we believe some of the strategies in Wellner and Zhan (1997) and Tsodikov (2003) may be needed in attacking this problem.

We note that, although sieve approximation was introduced to handle the situations where the ordinary likelihood does not work due to large nuisance parameter space (see Fan and Wong (2000)for a brief discussion), the sieve approximation in Section 6 is motivated by computational concerns. It seems desirable to develop an asymptotic theory for sieve profile likelihoods, which can be obtained by modifying the current theory, so as to get a good suggestion on the choice of sieves for computational purposes, and to see if the sieve profile likelihood is different from the theory in the present paper.

References

Bickel, P., Klaassen, C., Ritov, Y. and Wellner, J. A. (1993). *fficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins Univ. Press, Baltimore.

Bickel, P. and Ritov, Y. (2000). Comment on "On profile likelihood". *J. Amer. Statist. Assoc.* **95**, 466-468.

Chang, I. S., Hsiung, C. A., and Wen, C. C. (2002). An Asymptotic Theory for the Correlated Gamma-Frailty Model with Current Status Family Data. Technical Report DBB-2002-1, Division of Biostatistics and Bioinformatics, National Health Research Institutes, Taipei, Taiwan.

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

Dai, Y. H. and Yuan, Y. (1999). A nonlinear conjugate gradient method with a strong global convergence property. *SIAM J. Optim.* **10**, 177-182.

Fan, J. and Wong, W. H. (2000). Comment on "On profile likelihood". *J. Amer. Statist. Assoc.* **95**, 468-471.

Huang, J. (1996). Efficient estimation for the Cox model with interval censoring. *Ann. Statist.* **24**, 540-568.

Huang, J. and Wellner, J. A. (1995). Asymptotic normality of the NPMLE of linear functionals for interval censor data, case 1. *Statist. Neerlandica* **49**, 153-163.

Murphy, S. A. and van der Vaart, A. W. (1997). Semiparametric likelihood ratio inference. *Ann. Statist.* **25**, 1471-1509.

Murphy, S. A. and van der Vaart, A. W. (1999). Observed information in semiparametric models. *Bernoulli* **5**, 381-412.

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

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

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

Rudin, W. (1976). *Principles of Mathematical Analysis.* 3rd edition. McGraw-Hill, New York.

Stephen, H. F., Arnold, J. I. and Lawrence, E. S. (1997). *Linear Algebra.* 3rd edition. Prentice-Hall, Inc.

Tsodikov, A. (2003). Semiparametric models: a generalized self-consistency approach. *J. Roy. Statist. Soc. Ser. B* **65**, 3, 759-774.

Wellner, J. A. and Zhan, Y. (1997). A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *J. Amer. Statist. Assoc.* **92**, 945-959.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge Univ. Press, New York.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York. *Asymptotic Statistics.* Cambridge Univ. Press, New York.

Yashin, A., Vaupel, J. and Iachine, I. (1995). Correlated individual frailty: an advantageous approach to survival analysis of bivariate data. *Mathematical Population Studies* **5**, 145-159.

Institute of Cancer Research and Div. of Biostatistics and Bioinformatics, National Health Research Institutes, 35, Keyan Road, Zhunan Town, Miaoli County 350, Taiwan.

E-mail: ischang@nhri.org.tw

Department of Mathematics, Tamkang University, 151 Ying-chuan Road Tamsui, Taipei County 251, Taiwan.

E-mail: ccwen@mail.tku.edu.tw

Department of Applied Mathematics Chung Yuan Christian University, 200, Chung Pei Rd., Chung Li 320, Taiwan.

E-mail: yuhjenn@cycu.edu.tw

(Received February 2004; accepted January 2006)