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 $i$th individual in the $k$th 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
and at this time it is observed whether the survival time $T_{ik}$ has occurred or not, the observed data is \{(C_{ik}, \Delta_{ik}, 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 $X = (C, \Delta, Z_1, \ldots, C_m, \Delta_m, Z_m)$ denotes the data from a single family.

Let $H_0, H_1, \ldots, H_m$ be independent gamma distributed random variables with parameters $(\theta_1 \theta^{-2}, \theta^{-1}), (\theta_2 \theta^{-2}, \theta^{-1}), \ldots, (\theta_2 \theta^{-2}, \theta^{-1})$ respectively, where $(\theta_1, \theta_2) \in [0, \infty)^2 \setminus \{(0, 0)\}$, and $\theta = \theta_1 + \theta_2$. 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

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

where $\Lambda(\cdot)$ is a nondecreasing deterministic baseline function, and $\theta_3$ is a real number.

The quantity $\eta_0 + \eta_i$ in (1.1) is called the frailty for the $i$th individual, where $\eta_0$ is a common component for all individuals in the family and $\eta_i$ is an individual component. We note that the frailty variable $H_0 + H_i$ is gamma distributed with mean 1 and variance $\theta$. The correlation between $H_0 + H_i$ and $H_0 + H_j$, $i \neq j$, is $\theta_1 \theta^{-1}$. Hence, if $\theta_2$ 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 $\Lambda$. 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

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

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 $\eta$, we get the likelihood for $X$:

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

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

$$p(\eta; \theta_1, \theta_2) = \gamma(\eta_0; \theta_1 \theta^{-2}, \theta^{-1}) \prod_{i=1}^m \gamma(\eta_i; \theta_2 \theta^{-2}, \theta^{-1}).$$
The parameter space for \((\theta, \Lambda)\) we consider is \(\Theta \times \mathcal{L}\), where \(\Theta\) is a compact subset of \(((0, \infty)^2 \backslash \{(0, 0)\}) \times \mathbb{R}^1\) 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 \(\Lambda_0\) 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 \(\Theta\). This paper studies the inference problem regarding the frailty parameters \(\theta_1, \theta_2\), and the regression coefficient \(\theta_3\) 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, 1998, 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 $\theta$, and derives from it the asymptotic normality and efficiency of the NPMLE of $\theta$, the asymptotic distribution of the profile likelihood ratio statistic, and a consistent estimate of the covariance matrix of the NPMLE of $\theta$. 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 $\nu$, let $\nu^T$ denote its transpose, $\nu_i$ its $i$th 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

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

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

### 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 $(\theta_0, \Lambda_0)$. 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 $1/M < \Lambda_0(\tau_1) \leq \Lambda_0(\tau_2) < M$ 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

$$\left( \sum_{i=1}^3 u_i \frac{\partial}{\partial \theta_i} + \sum_{i=1}^m u_{i+3} \frac{\partial}{\partial y_i} \right)_{(\theta, y_1, \ldots, y_m) = (\theta_0, \Lambda_0(c^*_1), \ldots, \Lambda_0(c^*_m))}$$

$$\log \int p(\eta; \theta_1, \theta_2) \prod_{i=1}^m \left[ 1 - e^{-e^{\theta_3 z_i (\eta_0 + \eta_i) y_i}} \delta_i \left( e^{-e^{\theta_3 z_i (\eta_0 + \eta_i) y_i}} \right)^{1-\delta_i} d\eta = 0 \right. \quad (2.1)$$
Theorem 2.1. Identifiability of the parameters

2.1. Identifiability of the parameters defined on a sample space \( \Omega \) with a specific \( F \) or every positive integer \( n \).

Proof. Let

\[
\text{lik}(\theta, \Lambda) = \text{lik}(\theta_0, \Lambda_0) \text{ a.s. under } P_0, \text{ then } \theta = \theta_0 \text{ and } \Lambda = \Lambda_0 \text{ on } [\tau_1, \tau_2].
\]

**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 \( \theta \) at \((\theta_0, \Lambda_0)\) (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 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 \( \Omega \) with a specific \( \sigma \)-field.

2.1. Identifiability of the parameters

**Theorem 2.1.** There exists \( d^* > 0 \) such that if \( ||(\theta - \theta_0, (\Lambda - \Lambda_0)(c^*_i)) \rangle < d^* \), and \( \text{lik}(\theta, \Lambda) = \text{lik}(\theta_0, \Lambda_0) \text{ a.s. under } P_0, \text{ then } \theta = \theta_0 \text{ and } \Lambda = \Lambda_0 \text{ on } [\tau_1, \tau_2]. \)

**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

\[
P_0 \left( C_i \in [c_i, c_i + \frac{1}{n}), C_i \in [c^*_i, c^*_i + \frac{1}{n}) \text{ for } i = 2, \ldots, m; \Delta_i = 0, Z_i \in (z_i - \frac{1}{n}, z_i + \frac{1}{n}) \text{ for } i = 1, \ldots, m \right) > 0 \quad (2.2)
\]

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 \( \omega_n \) in \( \Omega \) 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 \( \text{lik}(\theta, \Lambda; X(\omega_n)) = \text{lik}(\theta_0, \Lambda_0; X(\omega_n)) \). Letting \( n \) go to infinity in \( \text{lik}(\theta, \Lambda; X(\omega_n)) = \text{lik}(\theta_0, \Lambda_0; X(\omega_n)) \), we obtain

\[
\int \text{p}(\eta; \theta_1, \theta_2) \exp(-e^{\theta^*_1}(\eta_0 + \eta_1)\Lambda(c_1)) \prod_{i=2}^m \exp(-e^{\theta^*_i}(\eta_0 + \eta_i)\Lambda(c^*_i)) d\eta
\]
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 \(i = 1, \ldots, m\) 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

\[
\int p(\eta; \theta_1, \theta_2) \prod_{i=1}^{m} \left[ 1 - e^{-\theta_0 z_i (\eta_0 + \eta_i)}\Lambda(c_i^*) \right]^{\delta_i} \left[ e^{-\theta_0 z_i (\eta_0 + \eta_i)}\Lambda(c_i^*) \right]^{1-\delta_i} d\eta
\]

= \[
\int p(\eta; \theta_1, \theta_2) \prod_{i=1}^{m} \left[ 1 - e^{-\theta_0 z_i (\eta_0 + \eta_i)}\Lambda_0(c_i^*) \right]^{\delta_i} \left[ e^{-\theta_0 z_i (\eta_0 + \eta_i)}\Lambda_0(c_i^*) \right]^{1-\delta_i} d\eta.
\]

Let \(G : \mathbb{R}^{m+3} \rightarrow \mathbb{R}^{m+3}\) be the vector valued function whose components are of the form

\[
(\theta, y_1, \ldots, y_m) \mapsto \log \int p(\eta; \theta_1, \theta_2) \prod_{i=1}^{m} \left[ 1 - e^{-\theta_0 z_i (\eta_0 + \eta_i)}y_i \right]^{\delta_i} \left[ e^{-\theta_0 z_i (\eta_0 + \eta_i)}y_i \right]^{1-\delta_i} d\eta.
\]

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 \(J_G(\theta_0, y_1, \ldots, y_m)\), 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 \(\mathbb{R}^{m+3}\) 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 \(\mathbb{R}^6\) to \(\mathbb{R}^6\) obtained from (2.1) by specifying \((\delta_1, \delta_2, \delta_3, z_1, z_2, z_3) = (1, 1, 1, 0, 0, 0), (1, 1, 0, 0, 0, 0), (1, 0, 1, 0, 1, 1), (0, 1, 0, 1, 0, 0)\) 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 \((\theta_0, \Lambda_0)\) on which the parameters are identifiable. Because the above determinant is an real analytic function of \((\theta_0, \Lambda_0(c_1^i), \ldots, \Lambda_0(c_m^i))\), 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 \(\{ (C_{ik}, \Delta_{ik}, Z_{1k}, \ldots, C_{mk}, \Delta_{mk}, Z_{mk}) \mid k = 1, \ldots, K \}\) is

\[
L_K(\theta, \Lambda) = \prod_{k=1}^{K} \int_{[0, \infty]^{m+1}} p(\gamma; \theta_1, \theta_2) \prod_{i=1}^{m} \left[ 1 - e^{e^{\theta_3} \Delta_{ik}(\eta_2 + \eta_i)\Lambda(C_{ik})} \right]^{\Delta_{ik}}
\]

\[
\times \left[ e^{e^{\theta_3} \Delta_{ik}(\eta_2 + \eta_i)\Lambda(C_{ik})} \right]^{1-\Delta_{ik}} d\eta.
\]

(2.4)

Since only the values of \(\Lambda\) at the \(C_{ik}\) matter in (2.4), all the estimates \(\hat{\Lambda}_K\) of \(\Lambda_0\) 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 \(\Lambda\) 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 \(\Lambda_0\), we take \(\hat{\gamma}_{ik} = \hat{\Lambda}_K(C_{ik})\) and \(\hat{Y}_{ik}\) to be the matrix \((\hat{\gamma}_{ik})\). Then the maximum likelihood estimate of \(\theta_0 = (\theta_{10}, \theta_{20}, \theta_{30})\) and \(\Lambda_0\) is the \(\hat{\theta}_K = (\hat{\theta}_1K, \hat{\theta}_2K, \hat{\theta}_3K)\) and \(\hat{\Lambda}_K\), represented by \(\hat{Y}_{ik}\) that maximizes

\[
\psi(\theta, Y) = \prod_{k=1}^{K} \int p(\gamma; \theta_1, \theta_2) \prod_{i=1}^{m} \left[ 1 - e^{e^{\theta_3} \Delta_{ik}(\eta_2 + \eta_i)\gamma_{ik}} \right]^{\Delta_{ik}} \left[ e^{e^{\theta_3} \Delta_{ik}(\eta_2 + \eta_i)\gamma_{ik}} \right]^{1-\Delta_{ik}} d\eta,
\]

subject to \((\theta, Y) \in \Theta \times D\), where

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

For fixed \(\theta\), 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 \( \psi \) are 1, and hence do not contribute anything to \( \psi \). 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  
\[
|\psi(\theta, Y)| \leq \int p(\eta; \theta_1, \theta_2) \exp(-e^{\theta_1 Z_{(mK)}(\eta_0 + \eta_1)y_{(mK)})} d\eta
\]

there exists \( d_0 > 0 \) such that  
\[
\max_{y_{(mK)} \leq d_0} \psi(\theta, Y) > \sup_{y_{(mK)} > d_0} \psi(\theta, Y).
\]  
(2.5)

Because \( \psi \) 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 \( \psi \) is uniformly continuous on \( \Theta \times D \), and hence the mapping \( \theta \mapsto L_K(\theta, \hat{\Lambda}_K(\cdot, \theta)) \) is continuous. Since \( \Theta \) is compact, the maximizer \( \hat{\theta}_K \) exists. Let \( \Lambda_K(\cdot) = \Lambda_K(\cdot, \hat{\theta}_K) \). 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  
\[
l_{\theta}(\theta, \Lambda)(X) = \left( l_{\theta_1}(\theta, \Lambda)(X), l_{\theta_2}(\theta, \Lambda)(X), l_{\theta_3}(\theta, \Lambda)(X) \right),
\]

where \( l_{\theta_i}(\theta, \Lambda)(X) = \frac{\partial}{\partial \theta_i} \log \lik(\theta; X) \). \( l_{\theta_i}(\theta, \Lambda) \) and \( l_{\theta}(\theta, \Lambda) \) are the score function for \( \theta_i \) and \( \theta \) at \( (\theta, \Lambda) \), respectively.
Let $\Lambda_{\varepsilon} = \Lambda + \varepsilon h$, where $\varepsilon$ is positive and $h$ is a nondecreasing and nonnegative function defined on $[0, \tau]$. The score function for $\Lambda$ in the direction $h$ at $(\theta, \Lambda)$, denoted by $l_\Lambda(\theta, \Lambda)[h]$, is defined by

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

A little calculation shows that

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

where

$$W_i(\theta, \Lambda; X) = \text{lik}(\theta, \Lambda; X)^{-1} \times \int p(\eta; \theta_1, \theta_2)q(\theta_3, \Lambda; \eta, X)e^{\theta_\Lambda Z_i(\eta_0 + \eta)}\frac{\Lambda_i}{1 - e^{\theta_\Lambda Z_i(\eta_0 + \eta)\Lambda(C_i)}} - 1 \ dn\eta.$$

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

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

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

$$h_j^*(u) = \left( A^{-1}\left(\sum_{i=1}^{m} E_0 \left[ l_{\theta_1}(\theta_0, \Lambda_0; X)W_i(\theta_0, \Lambda_0; X)|C_i = u \right] \right) \right)(u), \quad (3.3)$$

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

$$h_j^*(u) = \frac{E_0 \left[ l_{\theta_1}(\theta_0, \Lambda_0; X)\sum_{i=1}^{m} W_i(\theta_0, \Lambda_0; X)|C_1 = u \right]}{E_0 \left[ (\sum_{i=1}^{m} W_i(\theta_0, \Lambda_0; X))^2|C_1 = u \right]} \quad (3.4)$$

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 $l_\Lambda(\theta_0, \Lambda_0)[h^*] = (l_\Lambda(\theta_0, \Lambda_0)[h_1^*], l_\Lambda(\theta_0, \Lambda_0)[h_2^*], l_\Lambda(\theta_0, \Lambda_0)[h_3^*])$. 
Theorem 3.1. The efficient score function for θ at (θ₀, Λ₀) is \( \hat{l}_0 = l_\theta(θ₀, Λ₀) - l_Λ(θ₀, Λ₀)[h^*] \).

We note that the efficient Fisher information matrix for θ at (θ₀, Λ₀), denoted by \( I_0 \), is \( I_0 = P_0 \hat{l}_0^\top \hat{l}_0 \). We will show that \( I_0 \) is positive definite in Theorem 3.3.

Proof. It suffices to show that

\[
P_0(l_\theta(θ₀, Λ₀) - l_Λ(θ₀, Λ₀)[h^*_j])(l_Λ(θ₀, Λ₀)[h]) = 0,
\]

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

\[
E_0 \left[ l_\theta(θ₀, Λ₀; X) \sum_{i=1}^m h(C_i) W_i(θ₀, Λ₀; X) \right]
= E_0 \left[ \sum_{k=1}^m h^*(C_k) W_k(θ₀, Λ₀; X) \sum_{i=1}^m h(C_i) W_i(θ₀, Λ₀; X) \right].
\]

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

\[
\sum_{i=1}^m \int h(C_i) E_0 \left[ l_\theta(θ₀, Λ₀; X) W_i(θ₀, Λ₀; X) | C_i \right] dQ(C_i)
= \sum_{i=1}^m \int h(C_i) E_0 \left[ \sum_{k=1}^m h^*(C_k) W_k(θ₀, Λ₀; X) W_i(θ₀, Λ₀; X) | C_i \right] dQ(C_i).
\]

Thus (3.5) is equivalent to

\[
\int h(u) \sum_{i=1}^m E_0 \left[ l_\theta(θ₀, Λ₀; X) W_i(θ₀, Λ₀; X) | C_i = u \right] dQ(u)
= \int h(u) \sum_{i=1}^m E_0 \left[ \sum_{k=1}^m h^*(C_k) W_k(θ₀, Λ₀; X) W_i(θ₀, Λ₀; X) | C_i = u \right] dQ(u).
\]

In view of the definition of \( h^* \) 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 \( P_0(l_Λ(θ₀, Λ₀)[h])^2 = 0 \), and hence \( l_Λ(θ₀, Λ₀)[h] = 0 \) 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

\[
\begin{align*}
\int p(\eta; \theta_{10}, \theta_{20})e^{\theta_{30}z_i}(\eta_0 + \eta_i) \prod_{j=1}^{m} e^{-\theta_{30}z_j}(\eta_0 + \eta_j)A_0(c_1) \, d\eta_i = 0
\end{align*}
\]

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

\[
(Ah)(u) = h(u) \sum_{i=1}^{m} E_0(W_i^2(\theta_0, \Lambda_0; X)|C_i = u)
+ \sum_{i=1}^{m} \sum_{j=1}^{m} E_0(h(C_j)W_j(\theta_0, \Lambda_0; X)W_i(\theta_0, \Lambda_0; X)|C_i = u),
\]

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

\[
\|Ah - A_0h\|_{BV} = \left\| \sum_{i=1}^{m} \sum_{j=1}^{m} \int_{T_{2}} h(C_j)E_0(W_j(\theta_0, \Lambda_0; X)W_i(\theta_0, \Lambda_0; X)|C_i = u, C_j)dQ(C_j) \right\|_{BV}
\]

\[
\leq b\|h\|_{BV}, \quad (3.8)
\]

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 \( \nu = (\nu_1, \nu_2, \nu_3) \in \mathbb{R}^3 \). Since \( \nu I_0 \nu^T = P_0(\tilde{I}_0 \nu^T)^2 \geq 0 \), it suffices to show that \( \nu I_0 \nu^T = 0 \) implies \( \nu = 0 \).
Suppose $\nu_1T = 0$, then $\check{l}_0\nu^T = 0$ a.s. [P]. 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 $\check{l}_0\nu^T = 0$, and use the argument for deriving (2.3) to get

$$\left( \sum_{j=1}^3 \nu_j \frac{\partial}{\partial \theta_j} - \sum_{i=1}^m \left( \sum_{j=1}^3 \nu_j h_j^*(c_i^*) \frac{\partial}{\partial y_i} \right) \right) \log \int \frac{p(\eta; \theta_1, \theta_2)}{Q} \prod_{k=1}^m \left[ 1 - e^{-e^{\theta_3 y_k} (\eta_0 + \eta_k) y_k} \right]^{1-\delta_k} d\eta = 0$$

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 $(\hat{\theta}_K, \hat{\Lambda}_K) \in N_{\theta_0} \times L_\theta$, where $N_{\theta_0}$ is a neighborhood of $\theta_0$ and $L_\theta = \{ \Lambda = L_\theta | 1/M \leq \Lambda(\tau_1) \leq \Lambda(\tau_2) \leq M \}$. Define the profile likelihood for $\theta$,

$$pL_K(\theta) = \sup_{\Lambda \in L_\theta} L_K(\theta, \Lambda),$$

where $L_K(\theta, \Lambda)$ is the full likelihood given by (2.4). For every fixed $\theta$, 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 $\hat{\theta}_K \overset{P}{\to} \theta_0$,

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

where $\|\hat{\Lambda}_\theta - \Lambda_0\|_{2, Q} = (\int (\hat{\Lambda}_\theta - \Lambda_0)^2 dQ)^{1/2}$, 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 $\Psi = \{\log \text{lik}(\theta, \Lambda) \mid (\theta, \Lambda) \in \mathcal{N}_{\theta_0} \times \mathcal{L}_0\}$. For any probability measure $P$ on the sample space $\Omega$, 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

$$N_{\mathcal{C}}(\varepsilon, \mathcal{L}_2(P)) = \min \{N \mid \text{there exists } f_1^L, f_1^U, \ldots, f_N^L, f_N^U \text{ such that } \|f_n^L - f_n^U\|_{2,P} < \varepsilon, \text{ and for each } f \in \mathcal{C}, f_n^L \leq f \leq f_n^U \text{ for some } 1 \leq n \leq N\},$$

and the bracketing integral

$$J_{\mathcal{C}}(\varepsilon, \mathcal{L}_2(P)) = \int_0^\delta \sqrt{1 + \log N_{\mathcal{C}}(\varepsilon, \mathcal{L}_2(P))} \, d\varepsilon.$$

**Lemma 4.2.** $\log N_{\mathcal{C}}(\varepsilon, \mathcal{L}_2(P)) = O(1/\varepsilon)$ as $\varepsilon$ decreases to 0.

**Proof.** Because of the monotonicity of the elements in $\mathcal{L}_0$, we know $N_{\mathcal{C}}(\varepsilon, \mathcal{L}_0, L_2(Q)) \leq e^{c/\varepsilon}$ 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 $\{\Lambda_j^L, \Lambda_j^U, j = 1, \ldots, J\}$, where $J = O(e^{c/\varepsilon})$, such that $\|\Lambda_j^U - \Lambda_j^L\|_{2,Q} < \varepsilon$, and for each $\Lambda \in \mathcal{L}_0$, $\Lambda_j^L \leq \Lambda \leq \Lambda_j^U$ for some $1 \leq j \leq J$. Let $\Lambda_j^L = \Lambda_j^U - \varepsilon$ and $\Lambda_j^U = \Lambda_j^L + \varepsilon$. 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 $\varepsilon$ small enough that each $\Lambda_j^L$ also stays away from zero.

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

$$l_{j,\theta}^L(x) = \log \int p(\eta; \theta_1, \theta_2) \prod_{i=1}^m \left[1 - e^{-e^3x_i(\eta_0 + \eta_0\Lambda_j^L(c_i))}\right]^{\delta_i} \left[e^{-e^3x_i(\eta_0 + \eta_0\Lambda_j^U(c_i))}\right]^{1-\delta_i} \, d\eta,$$

$$l_{j,\theta}^U(x) = \log \int p(\eta; \theta_1, \theta_2) \prod_{i=1}^m \left[1 - e^{-e^3x_i(\eta_0 + \eta_0\Lambda_j^L(c_i))}\right]^{\delta_i} \left[e^{-e^3x_i(\eta_0 + \eta_0\Lambda_j^U(c_i))}\right]^{1-\delta_i} \, d\eta.$$

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

Consider the function

$$f(\alpha; x) = \log \int p(\eta; \theta_1, \theta_2) \prod_{i=1}^m \left[1 - e^{-e^3x_i(\eta_0 + \eta_0\eta_i)}\right]^{\delta_i} \left[e^{-e^3x_i(\eta_0 + \eta_0\eta_i)}\right]^{1-\delta_i} \, d\eta,$$

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

$$l_{j,\theta^\prime}^U(x) - \log \text{lik}(\theta, \Lambda; x)$$
\[ f(\alpha'; x) - f(\alpha''; x) = 3 \sum_{i=1}^{N} f_i(\tilde{\alpha})(\theta'_i - \theta_i) + \sum_{i=1}^{m} f_{i+3}(\tilde{\alpha}) \left[ \delta_i(\Lambda_j^U - \Lambda)(c_i) + (1 - \delta_i)(\Lambda_j^L - \Lambda)(c_i) \right]. \] (4.3)

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

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

for some constants \( b_0, b_1 > 0 \).

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

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

\[ I_{j,\theta^{(n)}}^{L} \leq \log \text{lik}(\theta, \Lambda) \leq I_{j,\theta^{(n)}}^{U}. \] (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

\[ \|I_{j,\theta^{(n)}}^{U} - I_{j,\theta^{(n)}}^{L}\|_2^2 \leq E \left( b_3 \sum_{i=1}^{m} |\Lambda_j^U(c_i) - \Lambda_j^L(c_i)|^2 \right) \]

\[ = b_3 m \|\Lambda_j^U - \Lambda_j^L\|_{L_2}^2 < b_3 m (3\varepsilon)^2 \] (4.6)

for some \( b_3 > 0 \). It follows from (4.5) and (4.6) that \( \mathcal{N}_{\{\varepsilon, \Psi, L_2(P)\}} \) is of the order \( NJ = O(e^{c}/\varepsilon^2) \), and hence \( \log \mathcal{N}_{\{\varepsilon, \Psi, L_2(P)\}} = O(1/\varepsilon) \). 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 \( (C_1, \ldots, C_m) = (c_1, \ldots, c_m) \),

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

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

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

**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

$$\dot{g}(\alpha_0^*; c^*_1, \delta_1, z_1,\ldots,c^*_m, \delta_m, z_m)u^T = 0$$

(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 (A.3). This completes the proof.

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

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

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

**Proof.** It suffices to show that

$$P_0\left(\log \text{lik}(\theta_0, \Lambda_0) - \log \text{lik}(\theta, \Lambda)\right) \leq b\|\theta - \theta_0\|^2,$$

(4.9)

$$P_0\left(\log \text{lik}(\theta_0, \Lambda_0) - \log \text{lik}(\theta, \Lambda_0)\right) \leq -b\|\Lambda - \Lambda_0\|_{2,Q}^2.$$  

(4.10)

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

Using the Taylor’s expansion of $g$ around $\alpha_0$, we know

$$P_0\left(\log \text{lik}(\theta, \Lambda) - \log \text{lik}(\theta_0, \Lambda_0)\right)$$

$$= E_0 E_0\left(g(\alpha; X) - g(\alpha_0; X) \mid C_1,\ldots,C_m\right)$$

$$= E_0\left(E_0(\ddot{g}(\alpha_0; X)|C_1,\ldots,C_m)(\alpha - \alpha_0)^T\right.$$

$$+ (\alpha - \alpha_0)\Sigma_{(c_1,\ldots,c_m)}(\alpha - \alpha_0)^T + o(\|\alpha - \alpha_0\|^2))$$

$$= E_0\left(l_0(\theta_0, \Lambda_0)(\theta - \theta_0)^T + l_\Lambda(\theta_0, \Lambda_0)[\Lambda - \Lambda_0](X)\right)$$
Recalling that $l_\theta$ and $l_\Lambda$ 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
\[
P_0 \left( \log \text{lik}(\theta, \Lambda) - \log \text{lik}(\theta_0, \Lambda_0) \right) \leq E_0 \left( \lambda(C_1, \ldots, C_m) \|\alpha - \alpha_0\|^2 + o(\|\alpha - \alpha_0\|^2) \right),
\]
where $\lambda(C_1, \ldots, C_m)$ is the largest eigenvalue of $\Sigma(C_1, \ldots, C_m)$. Noting that $\lambda$ is continuous at $(c_1^*, \ldots, c_m^*)$ and using Lemma 4.3, we know $\lambda$ 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
\[
E^* \sup_{\theta \in N_{\theta_0}, \Lambda \in L_0, \|\theta - \theta_0\| < \delta, \|\Lambda - \Lambda_0\|_2 < \delta} |G_K(\log \text{lik}(\theta, \Lambda) - \log \text{lik}(\theta_0, \Lambda_0))| \leq B \phi_K(\delta),
\]
for $\delta$ sufficiently small.

**Proof.** We first note that all elements of $\Psi$ are uniformly bounded. Using the Mean Value Theorem, there exists a constant $b > 0$ such that
\[
P_0 \left( \log \text{lik}(\theta, \Lambda) - \log \text{lik}(\theta_0, \Lambda_0) \right)^2 \leq b(\|\Lambda - \Lambda_0\|_2^2 + \|\theta - \theta_0\|^2)
\]
for every $(\theta, \Lambda) \in N_{\theta_0} \times L_0$.

Furthermore, we know from Lemma 4.2 that
\[
J(\delta, \Psi, L_2(P)) = O(\sqrt{\delta})
\]
as $\delta$ 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 $\|\hat{\Lambda}_{K^3} - \Lambda_0\|_2 = O_P(\|\hat{\theta}_K - \theta_0\| + K^{-1/3})$ for every random sequence $\hat{\theta}_K \rightarrow \theta_0$. This completes the proof.

### 5. Profile Likelihood Theory

In this section, we focus our attention on the estimation of $\theta$ 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 $\nu \mapsto (\nu, \Lambda_0 + h^*(\theta_0 - \nu)^T)$. The so-called least favorable submodel refers to this path. However, the second coordinate of the preceding path may not lie in the space $L_0$ defined in Section 4. We now modify and replace this path to obtain an approximately least-favorable submodel.

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

$$
\phi(y) = \begin{cases} 
0 & \text{for } 0 \leq y < \frac{1}{M}, \\
\frac{y - M^{-1}}{\Lambda_0(\tau_1) - M^{-1}} & \text{for } \frac{1}{M} \leq y < \Lambda_0(\tau_1), \\
1 & \text{for } \Lambda_0(\tau_1) \leq y < \Lambda_0(\tau_2), \\
\frac{M - y}{M - \Lambda_0(T_2)} & \text{for } \Lambda_0(\tau_2) \leq y \leq M.
\end{cases}
$$

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

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

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

$$u \mapsto \sum_{i=1}^m E_0(l_0(\theta_0, \Lambda_0; X)W_i(\theta_0, \Lambda_0; X)|C_i = u)$$

is Lipschitz. Let $\|\sum_i E_0(l_0(\theta_0, \Lambda_0; X)W_i(\theta_0, \Lambda_0; X)|C_i = \cdot )\|_{Lip} = L_0$, and let $\mathcal{H}_0$ be the closed linear span of $\{h \in \mathcal{H} \mid h \text{ is Lipschitz with } \|h\|_{Lip} \leq L_0\}$. We note that the proofs of Theorem 3.1 and Lemma 3.2 indicate that the operator $A : \mathcal{H}_0 \rightarrow \mathcal{H}_0$ 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.** $\Lambda_\nu(\theta, \Lambda)(\cdot)$ given by $(5.1)$ defines a cumulative hazard function in $L_0$ for every $\nu$ sufficiently close to $\theta$.

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

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

we know

$$\dot{l}(\theta_0, \theta_0, \Lambda_0; X) = \tilde{l}_0(X), \quad (5.3)$$
the efficient score function for $\theta$ at $(\theta_0, \Lambda_0)$. Furthermore, noting that $\nu \mapsto I(\nu, \theta_0, \Lambda_0)$ is a smooth parametric submodel, its derivatives at $\theta_0$ satisfies

$$P_0 \theta \tilde{I}(\theta_0, \theta_0, \Lambda_0) = -P_0 \theta \tilde{I}(\theta_0, \theta_0, \Lambda_0)\theta \tilde{I}(\theta_0, \theta_0, \Lambda_0) = -I_0.$$

(5.4)

**Lemma 5.2.** The class of functions $\{I(\nu, \theta, \Lambda) | (\nu, \theta, \Lambda) \in N_{\theta_0} \times N_{\theta_0} \times L_A\}$ is a uniformly bounded Donsker class, and the class of functions $\{I(\nu, \theta, \Lambda) | (\nu, \theta, \Lambda) \in N_{\theta_0} \times N_{\theta_0} \times L_A\}$ 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 $\theta_K \overset{P}{\to} \theta_0$,

$$P_0 \theta \tilde{I}(\theta_0, \theta_K, \Lambda_{\theta_K}) = o_P(||\theta_K - \theta_0|| + K^{-\frac{1}{2}}).$$

(5.5)

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

$$P_{\theta,\Lambda} \theta \tilde{I}(\theta_0, \theta, \Lambda) + P_{\theta,\Lambda} \partial \tilde{I}(\theta, \theta, \Lambda) \bigg|_{\nu = \theta} \bigg|_{\nu = \theta} = 0,$$

where $\tilde{I}(\theta, \theta, \Lambda)$ is the score function for $\theta$. Evaluating this at $(\theta, \Lambda) = (\theta_0, \Lambda_0)$ gives

$$-\frac{\partial}{\partial \nu} \bigg|_{\nu = \theta_0} P_{\theta,\Lambda} \theta \tilde{I}(\theta_0, \nu, \Lambda_0) = P_{\theta,\Lambda} \theta \tilde{I}(\theta_0, \theta_0, \Lambda_0) + P_{\theta,\Lambda} \theta \tilde{I}(\theta_0, \theta_0, \Lambda_0) = I_0 - I_0 = 0.$$

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

$$P_{\theta,\Lambda} \theta \tilde{I}(\theta_0, \theta, \Lambda) - \tilde{I}(\theta_0, \theta, \Lambda) = P_0 \bigg( \frac{\partial}{\partial \nu} \bigg|_{\nu = \theta_*} \tilde{I}(\theta_0, \nu, \Lambda) - \frac{\partial}{\partial \nu} \bigg|_{\nu = \theta_0} \tilde{I}(\theta_0, \nu, \Lambda_0) \bigg) (\theta - \theta_0)^T$$

for an intermediate point $\theta_*$ between $\theta$ and $\theta_0$. Letting $\theta = \hat{\theta}_K$ and $\Lambda = \hat{\Lambda}_{\theta_K}$, and using (4.1) and the Mean Value Theorem, we know it suffices to verify

$$P_{\theta,\Lambda} \theta \tilde{I}(\theta_0, \theta_0, \Lambda_{\theta_K}) = o_P(||\hat{\theta}_K - \theta_0|| + K^{-\frac{1}{2}})$$

(5.6)

to establish (5.5).

Noting that $P_{\theta_0,\Lambda} \theta \tilde{I}(\theta_0, \theta_0, \Lambda) = 0$, we have

$$P_0 \theta \tilde{I}(\theta_0, \theta_0, \Lambda) = (P_0 - P_{\theta_0,\Lambda})(\tilde{I}(\theta_0, \theta_0, \Lambda) - \tilde{I}(\theta_0, \theta_0, \Lambda_0)) + (P_0 - P_{\theta_0,\Lambda})\tilde{I}(\theta_0, \theta_0, \Lambda_0).$$

(5.7)
We now explain, without giving the details, that both terms on the right-hand side are bounded by a multiple of \( \| \Lambda - \Lambda_0 \|^2_{\mathbb{Q}} \). 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 \( (\Lambda(C_1), \ldots, \Lambda(C_m)) \rightarrow \text{lik}(\theta_0, \Lambda; X) \) around \( (\Lambda_0(C_1), \ldots, \Lambda_0(C_m)) \), and the fact \( l(\theta_0, \theta_0, \Lambda_0) \) is the efficient score \( I_0 \).

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

\textbf{Theorem 5.4.} For every random sequence \( \hat{\theta}_K \overset{P}{\rightarrow} \theta_0 \),

\[
\log p_L(K) - \log p_L(\theta_0) = (\hat{\theta}_K - \theta_0) \sum_{k=1}^{K} \overline{I}_0^2(X_k) - \frac{1}{2} K(\hat{\theta}_K - \theta_0) I_0(\hat{\theta}_K - \theta_0)^T
+ o_P(\sqrt{K} \| \hat{\theta}_K - \theta_0 \| + 1)^2.
\]

\textbf{Proof.} We apply Theorem 1 of \cite{Murphy2000} to prove this theorem. It is easy to see that the functions \( (\nu, \theta, \Lambda) \rightarrow l(\nu, \theta, \Lambda; X) \) and \( (\nu, \theta, \Lambda) \rightarrow \hat{l}(\nu, \theta, \Lambda; X) \) are continuous at \((\theta_0, \theta_0, \Lambda_0)\) for \( P_0 \)-almost every \( X \). For every random sequence \( \hat{\theta}_K \overset{P}{\rightarrow} \theta_0 \), (4.1) implies that \( \Lambda \hat{\theta}_K \overset{P}{\rightarrow} \Lambda_0 \). In view of (5.1), (5.3), (5.5) and Lemma 5.2, we know all conditions of Theorem 1 of \cite{Murphy2000} 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 \cite{Murphy2000}.

\textbf{Theorem 5.5.} The NPMLE \( \hat{\theta}_K \) is asymptotically normal and asymptotically efficient at \((\theta_0, \Lambda_0)\); that is,

\[
\sqrt{K}(\hat{\theta}_K - \theta_0) = I_0^{-1} \sqrt{K} P K \overline{I}_0^T + o_P(1) \overset{d}{\rightarrow} N(0, I_0^{-1}).
\]

\textbf{Theorem 5.6.} Under the null hypothesis \( H_0 : \theta = \theta_0 \), the profile likelihood ratio statistic

\[
lrt_K(\theta_0) \equiv 2 \log \frac{p_L(K)}{p_L(\theta_0)},
\]

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 $\nu_K \xrightarrow{P} \nu \in \mathbb{R}^3$ 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 $L_1 \subset L$, which comprises step functions with possible jump points $b_i$. A function $\Lambda$ in $L_1$ can thus be identified with a nonnegative vector $\mathbf{v} = (v_1, \ldots, v_N)$ with $v_j = \Lambda(b_j) - \Lambda(b_{j-1})$ for $j = 1, \ldots, N$; in fact, $\Lambda(c) = \sum_{j: b_j \leq c} v_j$. The sieve maximum likelihood estimates of $\theta_0$ and $\Lambda_0$ is the $\hat{\theta}$ and $\hat{\Lambda}$, represented by $\hat{\varphi}$, that maximizes

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

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 $\Phi$ relative to $\xi$ by $\Phi'$; 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 $\hat{\xi} = \xi^{(J)} + 2^{-n} \Phi'(\xi^{(J)})$.
5. If $\Phi(\xi^{(J)}) > \Phi(\hat{\xi})$, then set $n = n + 1$ and go back to (4).
6. If $\Phi(\xi^{(J)}) \leq \Phi(\hat{\xi})$, then set $\xi^{(J+1)} = \hat{\xi}$.
8. Repeat (3)~(7) for a suitable number $M$ of iterations once the evidence of convergence occurs.
9. Set $(\hat{\theta}, \hat{\varphi}) = (e^{\xi_1^{(M)}}, e^{\xi_2^{(M)}}, e^{\xi_3^{(M)}}, e^{\xi_4^{(M)}}, \ldots, e^{\xi_{N+3}^{(M)}})$.
We generate data with θ₀ = (1, 1, 0.5), Λ₀(t) = log(100/(100 - t)), Z₁ ∼ unif{0, 1}, C₁ ∼ unif(1, 99), m = 3, and K = 400. Our study consists of 100 replicates. The parameters for the sieve are N = 98 and bᵢ = i for i = 1, . . . , N. In applying the above algorithm, we take the starting value ξ(1) = (0, 0, 0.5, log(1/80), log(2/80), . . ., log(98/80)) to conduct the likelihood ratio test. Our simulation study seems to suggest that the χ² 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 H₀, and Figure 1 is the Q-Q plot for lrt₄₀₀(θ₀) versus χ²₃.

Table 1. Comparison of theoretical and simulated critical values (CV) and rejection rate (RR) under H₀.

<table>
<thead>
<tr>
<th>Significance Level (%)</th>
<th>CV for χ³</th>
<th>CV for lrt₄₀₀(θ₀)</th>
<th>RR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>0.5844</td>
<td>0.7058</td>
<td>93</td>
</tr>
<tr>
<td>80</td>
<td>1.0052</td>
<td>1.1144</td>
<td>82</td>
</tr>
<tr>
<td>70</td>
<td>1.4237</td>
<td>1.5589</td>
<td>72</td>
</tr>
<tr>
<td>60</td>
<td>1.8692</td>
<td>1.9569</td>
<td>61</td>
</tr>
<tr>
<td>50</td>
<td>2.3660</td>
<td>2.5947</td>
<td>51</td>
</tr>
<tr>
<td>40</td>
<td>2.9462</td>
<td>3.3848</td>
<td>47</td>
</tr>
<tr>
<td>30</td>
<td>3.6649</td>
<td>3.9840</td>
<td>36</td>
</tr>
<tr>
<td>20</td>
<td>4.6416</td>
<td>4.8706</td>
<td>21</td>
</tr>
<tr>
<td>10</td>
<td>6.2514</td>
<td>6.6243</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>7.8147</td>
<td>7.8293</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>11.3449</td>
<td>9.1968</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 1. Q-Q plot for lrt₄₀₀(θ₀) v.s. χ²₃.
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


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