On pseudolikelihood inference for semiparametric models with boundary problems

BY Y. CHEN

Department of Biostatistics and Epidemiology, University of Pennsylvania, 210 Blockley Hall, 423 Guardian Drive, Philadelphia, Pennsylvania 19104, U.S.A.
ychen123@upenn.edu

J. NING

Department of Biostatistics, The University of Texas MD Anderson Cancer Center, 1400 Pressler St, Unit 1411, FCT4.6006, Houston, Texas 77030, U.S.A.
jning@mdanderson.org

Y. NING

Department of Statistical Science, Cornell University, Comstock Hall 1188, Ithaca, New York 14853, U.S.A.
yn265@cornell.edu

K.-Y. LIANG

Department of Life Sciences, National Yang-Ming University, Taipei City 112, Taiwan
kyliang@ym.edu.tw

AND K. BANDEEN-ROCHE

Department of Biostatistics, Johns Hopkins University, 615 N. Wolfe St., Baltimore, Maryland 21205, U.S.A.
kbandeen@jhsphs.edu

SUMMARY

Consider a semiparametric model indexed by a Euclidean parameter of interest and an infinite-dimensional nuisance parameter. In many applications, pseudolikelihood provides a convenient way to infer the parameter of interest, where the nuisance parameter is replaced by a consistent estimator. The purpose of this paper is to establish the asymptotic behaviour of the pseudolikelihood ratio statistic under semiparametric models. In particular, we consider testing the hypothesis that the parameter of interest lies on the boundary of its parameter space. Under regularity conditions, we establish the equivalence between the asymptotic distributions of the pseudolikelihood ratio statistic and a likelihood ratio statistic for a normal mean problem with a misspecified covariance matrix. This result holds when the nuisance parameter is estimated at a rate slower than the usual rate in parametric models. We study three examples in which the asymptotic distributions are shown to be mixtures of
chi-squared variables. We conduct simulation studies to examine the finite-sample performance of the pseudolikelihood ratio test.

Some key words: Likelihood ratio test; Multivariate survival model; Pseudolikelihood; Semiparametric model.

1. Introduction

Consider a semiparametric model indexed by two parameters: a parameter of interest \( \theta \in \Omega \subset \mathbb{R}^d \) and a nuisance parameter \( \phi \) lying in a Banach space \( H \) with a norm \( \| \cdot \| \). Semiparametric models have been widely used in a variety of settings; see for example Gu & Zhang (1993), Murphy (1995), Huang (1996) and Cheng (2009). Asymptotic theory for semiparametric maximum likelihood estimation can be found in Bickel et al. (1993), van der Vaart (2000) and Kosorok (2008). The consistency of the bootstrap for M-estimators was established by Dixon et al. (2005) and Cheng & Huang (2010). Semiparametric likelihood ratio inference based on the profile likelihood has been developed by Murphy & van der Vaart (1997, 2000) and Banerjee (2005). Chen et al. (2014) studied the local identification of nonparametric and semiparametric models.

Pseudolikelihood provides an approach to inference on \( \theta \) in the presence of the nuisance parameter \( \phi \) (Gong & Samaniego, 1981). The key idea is that the inference for \( \theta \) can be based on \( L^*(\theta) = L(\theta, \hat{\phi}) \), where \( \hat{\phi} \) is a consistent estimator of \( \phi \) and \( L(\theta, \phi) \) is the loglikelihood. Unlike the profile likelihood, which estimates the nuisance parameter \( \phi \) by \( \hat{\phi}(\theta) = \operatorname{arg max}_{\phi \in H} L(\theta, \phi) \), the pseudolikelihood is constructed by substituting a consistent estimator \( \hat{\phi} \) that is free of \( \theta \), so the information equality does not hold (Gong & Samaniego, 1981). When the nuisance parameter \( \phi \) is of finite dimension, under certain regularity conditions Liang & Self (1996) derived the asymptotic distribution of the pseudolikelihood ratio test for \( \theta = \theta_0 \). One of the regularity conditions is that \( \theta_0 \) must be an interior point of its parameter space, but in many applications \( \theta_0 \) lies on the boundary of the parameter space. For parametric models, this boundary problem has been studied by Chernoff (1954), Kudo (1963), Chant (1974), Shapiro (1985), Self & Liang (1987) and Chen & Liang (2010). In particular, Chen & Liang (2010) derived the asymptotic distribution of the pseudolikelihood ratio test statistic for boundary problems when the nuisance parameter is of finite dimension. However, to the best of our knowledge, there is no systematic theoretical study of the boundary problem under semiparametric models.

The primary purpose of this paper is to develop a general theory on pseudolikelihood ratio inference for semiparametric models in cases where the parameter of interest may lie on the boundary of the parameter space. In a similar spirit to that of the profile likelihood (Murphy & van der Vaart, 2000), the theoretical justification for the pseudolikelihood in semiparametric models is more difficult than in Gong & Samaniego (1981), Liang & Self (1996) and Chen & Liang (2010), for the following reasons. First, the estimator of the nuisance parameter \( \phi \) may converge at a rate slower than the usual rate in parametric models. Second, unlike in parametric models, standard Taylor expansions cannot be used to deal with the remainder terms in likelihood expansions. To overcome these challenges, we establish our main results using empirical processes. Under certain regularity conditions, our main results cover cases in which the nuisance parameter is estimated with a rate slower than \( n^{1/2} \). In addition, the sensitivity of the likelihood to the nuisance parameter is characterized by the Fréchet derivative (Bickel et al., 1993). We establish a general theorem on the asymptotic distribution of the pseudolikelihood ratio test for \( \theta = \theta_0 \), which allows \( \theta_0 \) to lie on the boundary of its parameter space. The general theory is
verified and illustrated by copula and nested copula models for survival data and by weighted likelihoods for missing data. We show that the pseudolikelihood ratio test performs well in simulation studies, while the naive test that ignores the boundary problem has a conservative Type I error rate and much lower power.

2. Examples and main results

2.1. Examples

Example 1. Suppose that $C_\theta$ is a distribution function with density $c_\theta$ on $[0, 1]^2$ for some $\theta$ in $\mathbb{R}$. Let $(\tilde{Y}_1, \tilde{Y}_2)$ denote the paired failure times, and let $(S_1, S_2)$ and $(f_1, f_2)$ denote the corresponding marginal survival functions and density functions, respectively. Assuming that $(\tilde{Y}_1, \tilde{Y}_2)$ comes from the $C_\theta$ copula, the joint survival function and density function of $(\tilde{Y}_1, \tilde{Y}_2)$ are

$$S_\theta(y_1, y_2) = C_\theta[S_1(y_1), S_2(y_2)],$$

$$f_\theta(y_1, y_2) = c_\theta[S_1(y_1), S_2(y_2)]f_1(y_1)f_2(y_2) \quad (y_1, y_2 \geq 0).$$

Let $(C_1, C_2)$ denote paired censoring times. For $i = 1, \ldots, n$, assume that $(\tilde{Y}_{1i}, \tilde{Y}_{2i})$ and $(C_{1i}, C_{2i})$ are independent data. For each $i$, we observe $Y_{ji} = \tilde{Y}_{ji} \wedge C_{ji}$ and $\delta_{ji} = I(\tilde{Y}_{ji} \leq C_{ji})$ for $j = 1, 2$.

For concreteness, we consider the Clayton copula model (Clayton, 1978), defined by

$$C_\theta(u, v) = (u^{1-\theta} + v^{1-\theta} - 1)^{1/(1-\theta)},$$

for $\theta > 1$ and $C_\theta(u, v) = uv$ for $\theta = 1$. Given $n$ paired data $(Y_{1i}, Y_{2i}), \ldots, (Y_{ni}, Y_{mi})$, write $\{(S_1(Y_{1i}), S_2(Y_{2i}))\}$ as $(u_i, v_i)$ for notational simplicity. The loglikelihood function can be written as

$$L(\theta, S_1, S_2) = \sum_{i=1}^{n} \left[ \delta_{1i}\delta_{2i} \log c_\theta(u_i, v_i) + \delta_{1i}(1 - \delta_{2i}) \log \frac{\partial}{\partial u} C_\theta(u_i, v_i) \right]$$

$$+ (1 - \delta_{1i})\delta_{2i} \log \left[ \frac{\partial}{\partial v} C_\theta(u_i, v_i) \right] + (1 - \delta_{1i})(1 - \delta_{2i}) \log C_\theta(u_i, v_i),$$

where

$$\frac{\partial C_\theta(u, v)}{\partial u} = u^{-\theta}(u^{1-\theta} + v^{1-\theta} - 1)^{1/(1-\theta) - 1},$$

$$\frac{\partial C_\theta(u, v)}{\partial v} = v^{-\theta}(u^{1-\theta} + v^{1-\theta} - 1)^{1/(1-\theta) - 1},$$

$c_\theta(u, v) = \theta u^{-\theta} v^{-\theta} (u^{1-\theta} + v^{1-\theta} - 1)^{1/(1-\theta) - 2}$, and $\theta$ characterizes the association between the paired failure times $(\tilde{Y}_1, \tilde{Y}_2)$.

In practice, the marginal survival functions $S_1$ and $S_2$ are unknown nuisance parameters. To make inference on $\theta$, Shih & Louis (1995) proposed a log-pseudolikelihood function $L^*(\theta) = L(\theta, \hat{S}_1, \hat{S}_2)$, where $\hat{S}_1$ and $\hat{S}_2$ are Kaplan–Meier estimators of $S_1$ and $S_2$. In association analysis for bivariate survival times, a typical hypothesis of interest is $H_0 : \theta = 1$, i.e., no association between two failure times. The null hypothesis $\theta = 1$ lies on the boundary of the parameter.
space $\Omega = [1, \infty)$. However, this hypothesis testing problem with the boundary constraint is not covered by the theory of Shih & Louis (1995).

**Example 2.** Bandeen-Roche & Liang (1996) proposed a class of models for failure time data that accounts for multiple levels of clustering. For simplicity of notation, we consider a cluster of two levels, such as households and villages. Here, for illustration, we assume that there are three individual members and two households that are clustered as \{1\} and \{2, 3\}. When the Clayton copula is used to model the failure times ($\tilde{Y}_1, \tilde{Y}_2, \tilde{Y}_3$), its joint survival function is

$$S_\theta(y_1, y_2, y_3) = \left[ S_1(y_1)^{1-\theta_2} + \{ S_2(y_2)^{1-\theta_1} + S_3(y_3)^{1-\theta_1} - 1 \}^{(\theta_2-1)/(\theta_1-1)} - 1 \right]^{1/(1-\theta_2)},$$

where $(y_1, y_2, y_3) \in \mathbb{R}_+^3$ with $\mathbb{R}_+ = (0, \infty)$, $S_1(y_1), S_2(y_2)$ and $S_3(y_3)$ are the marginal survival functions, $\theta_1$ characterizes the association within the same household, and $\theta_2$ characterizes the association between two individuals from different households in the same village. To ensure that $S_\theta(y_1, y_2, y_3)$ is nonnegative, Bandeen-Roche & Liang (1996) required that $\theta_1 \geq \theta_2 \geq 1$.

Let $n$ denote the number of villages. For $i = 1, \ldots, n$, we observe $Y_{ji} = \tilde{Y}_{ji} \wedge C_{ji}$ and $\delta_{ji} = I(\tilde{Y}_{ji} \leq C_{ji})$ $(j = 1, 2, 3)$, where $(C_1, C_2, C_3)$ denote censoring times. Based on the observed data, we can specify the likelihood function, the form of which is provided in the Supplementary Material. Similar to Example 1, Bandeen-Roche & Liang (1996) proposed a pseudolikelihood approach for inference on $\theta$. One hypothesis of interest is $H_0 : \theta_1 = \theta_2 = 1$, i.e., no association among all failure times within the same village. In this example, $H_0$ is on the boundary of the parameter space $\Omega = \{(\theta_1, \theta_2) \in \mathbb{R}^2 : \theta_1 \geq \theta_2 \geq 1\}$.

**Example 3.** Assume that a study involves the collection of independent and identically distributed observations $(Y_i, X_i) (i = 1, \ldots, n)$, where $Y_i$ is the outcome and $X_i$ is a $d$-dimensional auxiliary covariate. Let $f(y; \theta)$ denote the density function of $Y_i$, which is indexed by a finite-dimensional parameter $\theta$. Let $V_i$ denote a binary missing data indicator, with $V_i = 1$ if $Y_i$ is observable and $V_i = 0$ if $Y_i$ is missing. Our goal is to estimate $\theta$ based on the observed data $(V_i Y_i, V_i, X_i) (i = 1, \ldots, n)$. Assume that $Y_i$ is independent of $V_i$ given $X_i$, known in the literature as missingness at random. Under this assumption, an inverse probability weighting estimator for $\theta$ is derived by solving a set of weighted estimating equations (Robins et al., 1994; Scharfstein et al., 1999). Under the likelihood framework, an equivalent estimator can be obtained by maximizing the weighted loglikelihood function

$$L(\theta, \pi) = \sum_{i=1}^{n} \frac{V_i}{\pi(X_i)} \log f(Y_i; \theta),$$

where $\pi(X_i) = \text{pr}(V_i = 1 \mid X_i)$ is the probability of observing $Y_i$ given covariates. The function $\pi(\cdot)$ is often an unknown infinite-dimensional nuisance parameter and must be estimated. When $X_i$ is low-dimensional, one can estimate $\pi$ by the Nadaraya–Watson estimator $\hat{\pi}$ (Nadaraya, 1964; Watson, 1964). The inference on $\theta$ can be based on the weighted pseudolikelihood function $L^*(\theta) = L(\theta, \hat{\pi})$. Suppose that $\theta$ is univariate and we are interested in the null hypothesis $H_0 : \theta = 0$. If there is prior knowledge that $\theta \geq 0$, then $\theta = 0$ is on the boundary of the parameter space $\Omega = [0, \infty)$.

### 2.2. Main results

In the following theoretical development, we consider a general setting. Given independent and identically distributed observations $(O_1, \ldots, O_n)$, let $L(\theta, \phi) = \sum_{i=1}^{n} m(\theta, \phi)(O_i)$ denote
Pseudolikelihood inference for semiparametric models

a generic objective function. For instance, \( L(\theta, \phi) \) reduces to the loglikelihood function if \( m(\theta, \phi)(O_i) \) is the log probability density function of the data \( O_i \); see Examples 1 and 2. The term \( m(\theta, \phi)(O_i) \) can also be a weighted loglikelihood function; see Example 3. Let \( \hat{\phi} \) denote an estimator of the nuisance parameter \( \phi \) in the Banach space \( \mathcal{H} \). Define \( L^*(\theta) = L(\theta, \phi) \). We refer to \( L^*(\theta) \) as the pseudolikelihood, even though \( L(\theta, \phi) \) can be an objective function other than the loglikelihood. The parameter space of \( \theta \) is denoted by \( \Omega \subset \mathbb{R}^d \). Let \( \hat{\theta} = \arg \max_{\theta \in \Omega} L^*(\theta) \) denote the maximum pseudolikelihood estimator. The pseudolikelihood ratio statistic \( T \) for the null hypothesis \( H_0 : \theta = \theta_0 \) versus the alternative hypothesis \( H_1 : \theta \in \Omega - \{\theta_0\} \) is defined as

\[
T = 2 \left\{ \sup_{\theta \in \Omega} L^*(\theta) - L^*(\theta_0) \right\}.
\]

To handle the nonparametric component \( \phi \), we introduce the following submodel notation (Kosorok, 2008). For any fixed \( \phi \), let \( \phi_t \in \mathcal{H} \) be a smooth curve running through \( \phi \) at \( t = 0 \). The loglikelihood for the parametric submodel indexed by \( (\theta, t) \) is \( L(\theta, \phi_t) \). Let \( a = (\partial / \partial t)\phi_t|_{t=0} \) be the direction vector in a tangent set for the nuisance parameter. Write

\[
m_1(\theta, \phi) = \frac{\partial}{\partial \theta} m(\theta, \phi), \quad m_2(\theta, \phi)[a] = \frac{\partial}{\partial t} m(\theta, \phi_t)|_{t=0},
\]

\[
m_{11}(\theta, \phi) = \frac{\partial^2}{\partial \theta \partial \theta^t} m(\theta, \phi), \quad m_{12}(\theta, \phi)[a] = \frac{\partial^2}{\partial \theta \partial t} m(\theta, \phi_t)|_{t=0}.
\]

Given a measurable function \( g \), we write \( \mathbb{P}_n g = n^{-1} \sum_{i=1}^n g(\omega_i) \) and write \( \mathbb{P} g = \int g \, d\mathbb{P} \) for the expectation of \( g \). We use \( | \cdot | \) to denote the \( L_2 \)-norm in the Euclidean space, and \( \| \cdot \| \) to denote the norm in the Banach space \( \mathcal{H} \). Let \( \phi_0 \) denote the true value of \( \phi \). We assume the following regularity conditions.

**Condition 1.** There exists some \( c_1 > 0 \) such that

\[|\hat{\theta} - \theta_0| = o_p(1), \quad \| \hat{\phi} - \phi_0 \| = O_p(n^{-c_1}).\]

**Condition 2.** For any \( \delta_n \to 0 \), any \( \theta \in \Omega \) and some \( D > 0 \),

\[\sup_{|\theta - \theta_0| \leq \delta_n, \| \phi - \phi_0 \| \leq D n^{-c_1}} |n^{1/2} (\mathbb{P}_n - \mathbb{P}) \{ m_1(\theta, \phi) - m_1(\theta_0, \phi_0) \} | = o_p(1).
\]

**Condition 3.** For some \( c_2 > 1 \) satisfying \( c_1 c_2 > 1/2 \) and any \( \| \phi - \phi_0 \| \leq D n^{-c_1} \),

\[\left| \mathbb{P} \{ m_1(\theta_0, \phi) - m_1(\theta_0, \phi_0) - m_12(\theta_0, \phi_0)(\phi - \phi_0) \} \right| = O(\| \phi - \phi_0 \|^2).
\]

**Condition 4.** As \( n \to \infty \), \( n^{1/2} \mathbb{P}_n m_1(\theta_0, \phi_0) \) and \( \mathbb{P} m_1(\theta_0, \phi_0) \{ n^{1/2} (\hat{\phi} - \phi_0) \} \) jointly converge in distribution to \( \mathcal{N}(0, \Sigma) \), where \( \Sigma \) is a positive-definite matrix and can be partitioned as \( \Sigma_{11}, \Sigma_{12} \) and \( \Sigma_{22} \) accordingly.

**Condition 5.** The information and covariance matrices

\[I_{11} = -\mathbb{P} m_{11}(\theta_0, \phi_0), \quad I_{11}^* = \Sigma_{11} + 2 \Sigma_{12} + \Sigma_{22}\]

are positive definite.
Condition 6. For any \( \delta_n \to 0 \), any \( \theta \in \Omega \) and some \( D > 0 \),

\[
\sup_{|\theta - \theta_0| \leq \delta_n, \|\phi - \phi_0\| \leq Dn^{-1}} \left| \mathbb{P}_n m_{11}(\theta, \phi) - \mathbb{P} m_{11}(\theta_0, \phi_0) \right| = o_P(1).
\]

A major difference between the above regularity conditions and those in the semiparametric literature is that we do not require \( \theta_0 \) to be interior to its parameter space. Moreover, in contrast to the efficient Fisher information matrix in profile likelihood estimation (Murphy & van der Vaart, 2000), for pseudolikelihood the information matrix \( I_{11}^* \) may not be identical to \( I_{11}^\ast \), where \( I_{11}^\ast \) is the covariance of the pseudo-score function \( n^{1/2} \mathbb{P}_n m_1(\theta_0, \phi_0) + \mathbb{P} m_{12}(\theta_0, \phi_0) \{ n^{1/2} (\hat{\phi} - \phi_0) \} \). This is also one of the major differences between the profile likelihood and the pseudolikelihood.

Our Conditions 1–6 are imposed on the model and the estimators instead of on the parameter space. These conditions are similar to the regularity conditions in the literature; see, for example, Cheng & Huang (2010). Specifically, Conditions 1–3 are similar to Conditions S1–S3 in Cheng & Huang (2010). Specifically, Conditions 1–3 are similar to Conditions S1–S3 in Cheng & Huang (2010). Specifically, Conditions 1–3 are similar to Conditions S1–S3 in Cheng & Huang (2010). Specifically, Conditions 1–3 are similar to Conditions S1–S3 in Cheng & Huang (2010). Specifically, Conditions 1–3 are similar to Conditions S1–S3 in Cheng & Huang (2010). Specifically, Conditions 1–3 are similar to Conditions S1–S3 in Cheng & Huang (2010). Specifically, Conditions 1–3 are similar to Conditions S1–S3 in Cheng & Huang (2010).
DEFINITION 1. The set $\Omega$ is approximated at $\theta_0$ by a cone with vertex at $\theta_0$, referred to as the approximating cone $C_\Omega(\theta_0)$, if

$$\inf_{x \in C_\Omega(\theta_0)} |x - y| = o(|y - \theta_0|), \quad y \in \Omega, \quad \inf_{y \in \Omega} |x - y| = o(|x - \theta_0|), \quad x \in C_\Omega(\theta_0).$$

For instance, if $\Omega = [1, \infty)$ and $\theta_0 = 1$, then $C_\Omega(\theta_0) = [1, \infty)$. Similarly, a sphere in $\mathbb{R}^2$ can be approximated at a boundary point by a half-plane tangent to the sphere at that point. In addition, if $\theta_0$ is an interior point of $\Omega \subset \mathbb{R}^d$, then $C_\Omega(\theta_0) = \mathbb{R}^d$. See Chernoff (1954), Shapiro (1985) and Self & Liang (1987) for more examples. Using Lemma 1, we derive the limiting distribution of the pseudolikelihood ratio statistic $T$ under the null hypothesis.

THEOREM 1. Assume that Conditions 1–6 hold and the parameter space $\Omega$ is approximated at $\theta_0$ by a cone $C_\Omega(\theta_0)$. Then the pseudolikelihood ratio statistic $T$ converges weakly to

$$T(Z) = Z^T I_{11} Z - \inf_{h \in C_\Omega(\theta_0)} \{ (Z - h)^T I_{11} (Z - h) \}$$

as $n \to \infty$, where $Z \sim N(0, I_{11}^{-1} I_{11}^{*} I_{11}^{-1})$ and $C_\Omega(0)$ is a cone obtained by translating $C_\Omega(\theta_0)$ so that its vertex lies at the origin. There are two scenarios as follows.

(i) If $\theta_0$ is an interior point of $\Omega$, then $T(Z)$ reduces to $Z^T I_{11} Z$ for $Z \sim N(0, I_{11}^{-1} I_{11}^{*} I_{11}^{-1})$, and $T(Z)$ is distributed as a weighted sum of $d$ independent $\chi^2_1$ variables with weights the eigenvalues of $I_{11}^{*} I_{11}^{-1}$.

(ii) If $\theta_0$ is a boundary point of $\Omega$, then the distribution of $T(Z)$ depends on the shape of $C_\Omega(0)$, and is generally a mixture of distributions as described by Chen & Liang (2010, Lemma 2).

A proof is given in the Appendix. This theorem shows that the asymptotic distribution of $T$ is the same as that of the likelihood ratio statistic for a normal mean problem with a misspecified covariance. To see why, consider a random variable $Z$ from the distribution $N(0, I_{11}^{-1} I_{11}^{*} I_{11}^{-1})$. If we incorrectly assume that $Z$ follows $N(h, I_{11}^{-1})$ with known covariance matrix $I_{11}^{-1}$ but unknown mean vector $h \in C_\Omega(0)$, the corresponding log-likelihood ratio test for testing $h = 0$ based on only one observation of $Z$ follows the same distribution as $T(Z)$.

The main difficulties when applying the general results in Theorem 1 are two-fold. First, for boundary problems the approximating cone $C_\Omega(\theta_0)$ has to be determined case by case. In § 2.3, we give its form and the calculation of $T(Z)$ for the examples of § 2.1. Further examples can be found in Self & Liang (1987) and Chen & Liang (2010). Second, the calculation of $I_{11}^{*}$ is more challenging than in Chen & Liang (2010) due to the presence of a nuisance parameter of infinite dimension. To address this issue, the following corollary provides an explicit formula for calculating $I_{11}^{*}$. In addition, this corollary deals with important special cases where the asymptotic distribution of the pseudolikelihood ratio statistic $T$ is simplified.

COROLLARY 1. Assume that the conditions of Theorem 1 hold and there exists a zero-mean function $\alpha(\theta_0, \phi_0)(O_i)$ such that

$$\mathbb{P} m_{12}(\theta_0, \phi_0) \{ n^{1/2} (\hat{\phi} - \phi_0) \} = n^{1/2} \mathbb{P} \alpha(\theta_0, \phi_0) + o_p(1).$$

Then, as $n \to \infty$, the asymptotic distribution of the pseudolikelihood ratio statistic $T$ is the same as the distribution of $T(Z)$ defined in (1), where $Z \sim N(0, I_{11}^{-1} I_{11}^{*} I_{11}^{-1})$ with

$$I_{11}^{*} = \text{cov} \{ m_1(\theta_0, \phi_0)(O_i) \} + 2 \text{cov} \{ m_1(\theta_0, \phi_0)(O_i), \alpha(\theta_0, \phi_0)(O_i) \} + \text{cov} \{ \alpha(\theta_0, \phi_0)(O_i) \}.$$
In addition, if \( P \{ 2m_1(\theta_0, \phi_0) + \alpha(\theta_0, \phi_0) \alpha^T(\theta_0, \phi_0) \} = 0 \) and \( P \{ m_1(\theta_0, \phi_0)m_1^T(\theta_0, \phi_0) \} = I_{11} \), we have \( I_{11}^* = I_{11} \) and therefore \( Z \sim N(0, I_{11}^{-1}) \).

The proof follows directly from Theorem 1. When \( I_{11}^* = I_{11} \), \( T \) has the same limiting distribution of the likelihood ratio statistic as it would have if the nuisance parameter were known, and the naive test that ignores the boundary constraints by comparing the pseudolikelihood ratio test statistic with \( \chi_d^2 \) always leads to conservative Type I error rates and a loss of power.

### 2.3. Examples revisited

**Example 1.** Recall that the null hypothesis \( \theta = 1 \) is on the boundary of the parameter space \( \Omega = [1, \infty) \). Then \( C_\Omega(0) = [0, \infty) \). Following the calculation in Shih & Louis (1995, p. 1389), the conditions in Corollary 1 hold and we obtain \( I_{11}^* = I_{11} \). Therefore, equation (1) reduces to

\[
T(Z) = Z^2I_{11} - Z^2I(Z < 0)I_{11} = Z^2I(Z > 0)I_{11},
\]

where \( Z \sim N(0, I_{11}^{-1}) \). Thus, the asymptotic distribution of \( T \) is a mixture of \( \chi_0^2 \) and \( \chi_1^2 \) with mixing probabilities 0.5 and 0.5, where \( \chi_0^2 \) is a point mass at 0. Additional details on the verification of Conditions 1–6 are given in the Supplementary Material.

**Example 2.** Denote the parameter value under the null hypothesis by \( \theta_0 = (1, 1)^T \). The approximating cone in this case is \( C_\Omega(\theta_0) = \{ (t_1, t_2) : t_1 \geq t_2 \geq 0 \} \). For ease of derivation, a simple reparameterization from \( \theta = (\theta_1, \theta_2)^T \) to \( \tau = (\tau_1, \tau_2)^T \), where \( \tau_1 = \theta_2 - 1 \) and \( \tau_2 = \theta_1 - \theta_2 \), yields the approximating cone \( C_\Omega(\tau_0) = [0, \infty) \times [0, \infty) \) with \( \tau_0 = (0, 0)^T \), which is illustrated in Fig. 1(a).

By arguments similar to those in Shih & Louis (1995), equation (1) reduces to

\[
T(Z_\tau) = Z_\tau^TI_{\tau\tau}Z_\tau - \inf_{\tau \in (0, \infty) \times (0, \infty)} (Z_\tau - \tau)^TI_{\tau\tau}(Z_\tau - \tau),
\]

where \( Z_\tau \sim N(0, I_{\tau\tau}^{-1}) \) and \( I_{\tau\tau} = E\{ -\partial^2 \log f(O_i; \tau_0, \phi_0) / \partial \tau^2 \} \). Here \( \phi_0 = (S_{10}, S_{20}, S_{30}) \) are the true values of \( (S_1, S_2, S_3) \), \( O_i = (Y_{i1}, Y_{i2}, Y_{i3}) \), and \( f(\cdot) \) is the density function of \( O_i \). Let

![Fig. 1. The partitions of parameter space considered in Example 2: (a) the parameter space for \( \tau = (\tau_1, \tau_2) \), with the shaded region representing admissible parameter values; (b) the transformed parameter space, where the shaded region \( \tilde{C}_\Omega(\tau_0) \) represents admissible parameter values. The asymptotic distribution of \( \tilde{T} \) is a mixture of \( \chi_0^2 \), \( \chi_1^2 \) and \( \chi_0^2 \) distributions, with mixing probabilities depending on the angles in \( \tilde{C}_\Omega(\tau_0) \).](https://academic.oup.com/biomet/article-abstract/104/1/165/3003354/fig1)
Pseudolikelihood inference for semiparametric models

\( I_{\tau\tau} = R^T R \), where \( R \) is a \( 2 \times 2 \) nonsingular matrix, and write \( \tilde{C}_\Omega(\tau_0) = \{ \tilde{\tau} : \tilde{\tau} = R \tau \) for any \( \tau \in C_\Omega(\tau_0) \) and \( \tilde{Z}_\tau = RZ_\tau \). Then \( T(Z_\tau) \) can be rewritten as

\[
T(\tilde{Z}_\tau) = |\tilde{Z}_\tau|^2 - \inf_{\tilde{\tau} \in \tilde{C}_\Omega(\tau_0)} |\tilde{Z}_\tau - \tilde{\tau}|^2.
\]

The calculation of the second term in (2) depends on the location of \( \tilde{Z}_\tau \) relative to the boundary of \( \tilde{C}_\Omega(\tau_0) \). Four different regions must be considered separately, as illustrated in Fig. 1(b): the shaded region represents \( \tilde{C}_\Omega(\tau_0) \); the angle in the shaded area is less than 180°, since the convexity of \( C_\Omega(\tau_0) \) is preserved under the linear transformation \( \tau \rightarrow R \tau \).

Denote the columns of \( R \) by \( R_1 \) and \( R_2 \), and denote the inner product of vectors \( a \) and \( b \) by \( \langle a, b \rangle = a^T b \). Then (2), namely the asymptotic distribution of \( T \), can be written as

\[
T(\tilde{Z}_\tau) = \begin{cases} 
|\tilde{Z}_\tau|^2 \sim \chi^2_2, & \tilde{Z}_\tau \in \tilde{C}_\Omega(\tau_0), \\
\left(\frac{\langle \tilde{Z}_\tau , R_2 \rangle}{|R_2|}\right)^2 \sim \chi^2_1, & \tilde{Z}_\tau \in \text{region 1,} \\
\left(\frac{\langle \tilde{Z}_\tau , R_1 \rangle}{|R_1|}\right)^2 \sim \chi^2_1, & \tilde{Z}_\tau \in \text{region 2,} \\
0, & \tilde{Z}_\tau \in \text{region 3.}
\end{cases}
\]

Since the distribution of \( \tilde{Z}_\tau \) is symmetric about the origin, the probabilities of \( \tilde{Z}_\tau \) being from certain regions are completely determined by the angles of these regions (Chernoff, 1954). The mixing probability for the shaded region is

\[
p_s = \cos^{-1}\left[\left\{I^{(1,1)}_{\tilde{\tau}} I^{(2,2)}_{\tilde{\tau}} \right\}^{-1/2} I^{(1,2)}_{\tilde{\tau}}\right] / (2\pi),
\]

where \( I^{(i,j)}_{\tilde{\tau}} \) is the \((i,j)\) element of the \( 2 \times 2 \) matrix \( I_{\tilde{\tau}\tilde{\tau}} \). Therefore, under \( H_0 : \theta_1 = \theta_2 = 1 \), the asymptotic distribution of the pseudolikelihood ratio test \( T \) is a mixture of \( \chi^2_2 \), \( \chi^2_1 \) and \( \chi^0_0 \) distributions with mixing probabilities \( p_s \), 0.5 and 0.5 - \( p_s \), respectively.

**Example 3.** We consider the cases \( \Omega = [0, \infty) \) and \( \Omega = \mathbb{R} \). In the former, the approximating cone is \( C_\Omega(0) = [0, \infty) \). So (1) reduces to \( T(Z) = Z^2 I_{11} - Z^2 I(Z < 0) I_{11} = Z^2 I(Z > 0) I_{11} \), where \( Z \sim N(0, I_{11}^{-1} I_{11}^{-1} \chi^2_1) \). The asymptotic distribution of \( T \) is a mixture of \( \chi^2_0 \) and \( I_{11}^{-1} \chi^2_1 \) distributions with mixing probabilities 0.5 and 0.5. In the latter case, \( \theta = 0 \) is an interior point and \( C_\Omega(0) = \mathbb{R} \). Hence, the asymptotic distribution of \( T \) is weighted chi-squared, \( I_{11}^{-1} \chi^2_1 \). Unlike the previous two examples, in which the nuisance parameter is estimated at an \( n^{1/2} \) rate, the Nadaraya–Watson estimator \( \hat{\tau} \) in this example has a slower rate. The explicit forms of \( I^{\ast}_{11} \) and \( I_{11} \) and details on verifying Conditions 1–6 are presented in the Supplementary Material.

3. Simulations

We conducted simulation studies in the settings of Examples 1 and 2. We first applied the pseudolikelihood ratio test for the dependence between bivariate survival times in Example 1. To generate the paired failure times, we used the rmvdc function in the R package copula (Yan, 2007; R Development Core Team, 2017). The marginal distributions of the failure times were
At the nominal level of 5%, the pseudolikelihood ratio test had about 80% power when generated from uniform distributions on \((0, 5-4)\) and \((0, 2-7)\), corresponding to 15% and 30% censoring. To evaluate the size of the test, we drew bivariate failure times from a Clayton copula with \(\theta = 1\), corresponding to an independence scenario. To evaluate the power of the tests, we implemented a similar procedure but set \(\theta\) to 1.3, 1.5, 1.7 and 2.0. We set the number of pairs at 100, 200 or 400. For each generated dataset, we compared the pseudolikelihood ratio statistic with the \(0.5\chi_0^2 + 0.5\chi_1^2\) distribution. Table 1 shows the estimated levels of Type I error and power from 5000 replications of the test. When the null hypothesis is true, the rejection rates of the pseudolikelihood ratio test were all within 95% confidence intervals for the nominal levels, i.e., 0.7–1.3% for nominal level 1% and 4.4–5.6% for nominal level 5%. A plot of the quantiles of the test statistics against those of the asymptotic distribution indicates that the latter works well at levels other than 5% and 1%; see the Supplementary Material. We also compared the test statistics with a \(\chi_1^2\) distribution as if the boundary constraint were ignored. This naive test was too conservative under all scenarios considered; see the Supplementary Material. The power of the pseudolikelihood ratio test increased with increasing values of the association parameter \(\theta\).

Weibull with shape parameter 2 and unit scale parameter. Independent censoring times were generated from uniform distributions on \((0, 5-4)\) and \((0, 2-7)\), corresponding to 15% and 30% censoring. To evaluate the size of the test, we drew bivariate failure times from a Clayton copula with \(\theta = 1\), corresponding to an independence scenario. To evaluate the power of the tests, we implemented a similar procedure but set \(\theta\) to 1.3, 1.5, 1.7 and 2.0. We set the number of pairs at 100, 200 or 400. For each generated dataset, we compared the pseudolikelihood ratio statistic with the \(0.5\chi_0^2 + 0.5\chi_1^2\) distribution. Table 1 shows the estimated levels of Type I error and power from 5000 replications of the test. When the null hypothesis is true, the rejection rates of the pseudolikelihood ratio test were all within 95% confidence intervals for the nominal levels, i.e., 0.7–1.3% for nominal level 1% and 4.4–5.6% for nominal level 5%. A plot of the quantiles of the test statistics against those of the asymptotic distribution indicates that the latter works well at levels other than 5% and 1%; see the Supplementary Material. We also compared the test statistics with a \(\chi_1^2\) distribution as if the boundary constraint were ignored. This naive test was too conservative under all scenarios considered; see the Supplementary Material. The power of the pseudolikelihood ratio test increased with increasing values of the association parameter \(\theta\).

At the nominal level of 5%, the pseudolikelihood ratio test had about 80% power when \(\theta\) was 1.7 at a sample size of 100, when \(\theta\) was 1.5 at a sample size of 200, and when \(\theta\) was 1.3 at a sample size of 400. The power slightly decreased with increased censoring; thus censoring has a relatively small effect on the power.

In the second simulation study, we tested for associations among all failure times within the same village in the model of Example 2. To generate the multivariate failure times with the hierarchical structure, we used the R package nacopula (Hofert & Mächler, 2011). The marginal distributions of the failure times were chosen to be standard exponential. Independent censoring times were generated from uniform distributions to have censoring percentages of 15% and 30%. For each generated dataset, we compared the pseudolikelihood ratio test statistic with the \((0.5 - \hat{p}_s)\chi_0^2 + 0.5\chi_1^2 + \hat{p}_s\chi_2^2\) distribution, where \(\hat{p}_s\) can be estimated empirically by (3), as described in § 2.3. Table 2 shows the estimated Type I error rates and power from 5000

### Table 1. Empirical rejection rates (%) of the pseudolikelihood ratio test for testing association between bivariate survival times in Example 1 over 5000 replications

| \(n\) | \(\theta\) | Censoring % = 0% | Censoring % = 15% | Censoring % = 30% |
| --- | --- | --- | --- | --- |
| | | Rejection | Rejection | Rejection | Rejection | Rejection | Rejection |
| 100 | 1.0 | 5.5 | 1.3 | 5.4 | 1.1 | 5.4 | 1.1 |
| | 1.3 | 36.8 | 16.6 | 34.6 | 14.6 | 33.5 | 14.1 |
| | 1.5 | 65.2 | 39.3 | 62.9 | 36.5 | 60.2 | 34.2 |
| | 1.7 | 84.4 | 65.0 | 82.3 | 61.2 | 80.2 | 58.4 |
| | 2.0 | 97.1 | 88.6 | 95.9 | 86.1 | 94.7 | 83.7 |
| 200 | 1.0 | 5.4 | 1.0 | 5.5 | 1.0 | 5.2 | 1.1 |
| | 1.3 | 55.8 | 29.6 | 55.5 | 29.5 | 55.5 | 28.6 |
| | 1.5 | 87.0 | 68.6 | 86.3 | 67.4 | 85.5 | 66.7 |
| | 1.7 | 97.7 | 91.6 | 97.4 | 90.1 | 96.8 | 89.3 |
| | 2.0 | 99.9 | 99.5 | 99.9 | 99.1 | 99.8 | 98.7 |
| 400 | 1.0 | 5.2 | 1.1 | 5.1 | 1.0 | 5.4 | 1.3 |
| | 1.3 | 77.2 | 53.7 | 76.0 | 52.8 | 76.1 | 52.2 |
| | 1.5 | 98.6 | 93.3 | 98.1 | 91.6 | 97.8 | 91.6 |
| | 1.7 | 100.0 | 99.8 | 99.9 | 99.6 | 99.9 | 99.4 |
| | 2.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
4. Discussion

If the maximizer of the log-pseudolikelihood $L^*(\theta)$ does not have a closed-form solution, iterative algorithms are needed to maximize $L^*(\theta)$. Cheng (2013) provided a general algorithm for maximizing the log profile likelihood and established its rate of convergence. Unlike the algorithm in Cheng (2013), the pseudolikelihood approach does not require iterative updating of the nuisance parameter estimate, since $\hat{\phi}$ is free of $\theta$ by definition. It is of interest to extend the theoretical results on the algorithm in Cheng (2013) to pseudolikelihood estimation.

In regular statistical models, the likelihood ratio test is known to be asymptotically optimal, whereas the pseudolikelihood ratio test may lose efficiency due to the use of a generic estimator of the nuisance parameter. For a class of nonregular models in which the parameters are not identifiable under the null hypothesis, Song et al. (2009) proposed optimal tests based on the integrated profile likelihood. One future direction of research is to study the optimality of these tests when the parameter of interest lies on the boundary of the parameter space.

This paper focuses on the pseudolikelihood, which relies on the availability of a consistent estimator for the nuisance parameter that is free of the parameter of interest. In some situations, such a consistent estimator may not be available. In such cases, likelihood ratio inference can be considered. The theoretical results for the semiparametric likelihood ratio test simulations. Similar to our findings in Example 1, the proposed test had sizes close to nominal, suggesting that the asymptotic approximation performs well. The naive test, which ignores the boundary problem, led to conservative Type I error rates and substantial loss of power; see the Supplementary Material.
Similar to that given in the present paper. We leave this work for future investigation.

Acknowledgement

We thank the editor, the associate editor and three referees for their helpful comments, which have significantly improved the paper. The authors are grateful to Drs John Breitner and Peter Zandi for providing the Cache County data. This research was supported in part by the U.S. National Institutes of Health.

Supplementary material

Supplementary material available at Biometrika online includes more technical details of Examples 1–3, further results from the simulation studies, and a real-data example.

Appendix

Proof of Lemma 1

We apply a second-order Taylor expansion to the pseudolikelihood ratio:

\[
2n P_a\{m(\theta, \hat{\phi}) - m(\theta_0, \hat{\phi})\} = 2n^{1/2}\{n^{1/2}(\theta - \theta_0)\}^T\bar{P}_a m_1(\theta_0, \hat{\phi}) + n^{1/2}(\theta - \theta_0)^T\bar{P}_a m_1(\hat{\theta}, \hat{\phi})n^{1/2}(\theta - \theta_0),
\]

where \(\hat{\theta} = \theta_0 + t(\theta - \theta_0)\) for some \(t \in [0, 1]\). By the convexity of \(\Omega\), we have \(\hat{\theta} \in \Omega\). By Condition 2,

\[
n^{1/2}(\bar{P}_a - \bar{P})m_1(\theta_0, \hat{\phi}) = n^{1/2}(\bar{P}_a - \bar{P})m_1(\theta_0, \phi_0) + o_p(1).
\]

Using \(\bar{P}_a m_1(\theta_0, \phi_0) = 0\) together with Condition 3, we obtain

\[
n^{1/2}\bar{P}_a m_1(\theta_0, \hat{\phi}) = n^{1/2}\bar{P}_a m_1(\theta_0, \hat{\phi}) + n^{1/2}\bar{P}_a m_1(\theta_0, \phi_0) + o_p(1) = n^{1/2}\bar{P}_a m_1(\theta_0, \phi_0) + \bar{P}_a m_1(\theta_0, \phi_0)\{n^{1/2}(\hat{\phi} - \phi_0)\}
\]

\[
+ O_p(n^{1/2}\|\hat{\phi} - \phi_0\|^2) + o_p(1).
\]

Since \(\|\hat{\phi} - \phi_0\| = O_p(n^{-c_1})\) and \(c_1c_2 > 1/2\), we have \(n^{1/2}\|\hat{\phi} - \phi_0\|^2 = o_p(1)\). By Conditions 5 and 6,

\[
\bar{P}_a m_1(\hat{\theta}, \hat{\phi}) = -I_{11} + o_p(1).
\]

Combining (A1), (A2) and (A3), we obtain the following quadratic expansion of the likelihood ratio statistic:

\[
2n\bar{P}_a\{m(\theta, \hat{\phi}) - m(\theta_0, \hat{\phi})\} = 2\{n^{1/2}(\theta - \theta_0)\}^Tn^{1/2}\bar{P}_a m_1(\theta_0, \phi_0)
\]

\[
+ 2\{n^{1/2}(\theta - \theta_0)\}^T\bar{P}_a m_1(\theta_0, \phi_0)\{n^{1/2}(\hat{\phi} - \phi_0)\}
\]

\[
- n^{1/2}(\theta - \theta_0)^TI_{11}n^{1/2}(\theta - \theta_0) + o_p(1 + n^{1/2}\|\theta - \theta_0\|^2),
\]

\[\text{(A4)}\]
which is equivalent to

\[ 2n\mathbb{P}_n\{m(\hat{\theta}, \hat{\phi}) - m(\theta_0, \phi_0)\} = -W(\theta)^T I_{11} W(\theta) + U(\theta_0)^T I_{11}^{-1} U(\theta_0) + o_p(1 + n^{1/2}\|\theta - \theta_0\|)^2 \]  

(A5)

where \( W(\theta) = -n^{1/2}(\theta - \theta_0) + I_{11}^{-1} U(\theta_0) \) and \( U(\theta_0) = n^{1/2}\mathbb{P}_n m_1(\theta_0, \phi_0) + \mathbb{P} m_{12}(\theta_0, \phi_0) \{n^{1/2}(\hat{\phi} - \phi_0)\} \).

This completes the proof.

**Proof of Theorem 1**

For notational simplicity, let \( \hat{\theta} = n^{1/2}(\hat{\theta} - \theta_0) \) and \( U = n^{1/2}\mathbb{P}_n m_1(\theta_0, \phi_0) + \mathbb{P} m_{12}(\theta_0, \phi_0) \{n^{1/2}(\hat{\phi} - \phi_0)\} \). Replacing \( \theta \) in (A4) by \( \hat{\theta} \), we have

\[ 2n\mathbb{P}_n\{m(\hat{\theta}, \hat{\phi}) - m(\theta_0, \phi_0)\} = 2\hat{h}^T U - \hat{h}^T I_{11} \hat{h} + o_p(1 + |\hat{h}|)^2. \]

Since \( \mathbb{P}_n\{m(\hat{\theta}, \hat{\phi}) - m(\theta_0, \phi_0)\} \geq 0 \), \( U \) is bounded in probability by Condition 4 and \( I_{11} \) is positive definite by Condition 5, we have

\[ 0 \leq K|\hat{h}| - K'|\hat{h}|^2 + o_p(1 + |\hat{h}|)^2 \]

for some positive constants \( K \) and \( K' \). If \( |\hat{h}| = o_p(1) \), then \( |\hat{h}| = O_p(1) \). Otherwise, \( 1 + |\hat{h}| \) is of the order of \( |\hat{h}| \), and we have \( K'|\hat{h}|^2 \leq K|\hat{h}| + o_p(|\hat{h}|^3) \). This implies that \( |\hat{h}| = O_p(1) \) and thus \( \hat{h} \) is uniformly tight. Write \( W_h = -\hat{h} + I_{11}^{-1} U \). In the following, we shall prove that

\[ 2n\mathbb{P}_n\{m(\hat{\theta}, \hat{\phi}) - m(\theta_0, \phi_0)\} = \sup_{h \in C\Omega_1} (-W_h^T I_{11} W_h + U^T I_{11}^{-1} U) = o_p(1). \]  

(A6)

By Condition 5, we have that \( U \) converges weakly to \( N(0, I_{11}^{-1}) \). By (A6) and the continuous mapping theorem,

\[ \sup_{h \in C\Omega_1} (-W_h^T I_{11} W_h + U^T I_{11}^{-1} U) \rightarrow \sup_{h \in C\Omega_1} \{-Z^2 + Z^2 I_{11} I_{11} Z\} \]

as \( n \to \infty \), where \( Z \sim N(0, I_{11}^{-1}) \). Hence, Slutsky’s theorem implies that the pseudolikelihood ratio test \( 2n\mathbb{P}_n\{m(\hat{\theta}, \hat{\phi}) - m(\theta_0, \phi_0)\} \) converges weakly to \( \sup_{h \in C\Omega_1} \{-Z^2 + Z^2 I_{11} I_{11} Z\} \)

It remains to show that (A6) holds. Since \( \hat{\theta} \) is root-\( n \)-consistent, (A5) gives

\[ 2n\mathbb{P}_n\{m(\hat{\theta}, \hat{\phi}) - m(\theta_0, \phi_0)\} = \sup_{h \in \Omega_n} (-W_h^T I_{11} W_h + U^T I_{11}^{-1} U) = o_p(1), \]

where \( \Omega_n = \{n^{1/2}(\theta - \theta_0) : \theta \in \Omega\} \). Comparing with (A6), we only need to show that \( \inf_{h \in \Omega_n} W_h^T I_{11} W_h = \inf_{h \in C\Omega_1} W_h^T I_{11} W_h + o_p(1) \). Similar to the proof of root-\( n \) convergence of \( \hat{\theta} \), we can show that the minimizer of \( W_h^T I_{11} W_h \) in \( \Omega_n \) is bounded in probability. By the definition of \( \Omega_n \), for any \( h \in \Omega_n \) with \( |h| = O(1) \), there exists \( \theta \in \Omega \) such that \( h = n^{1/2}(\theta - \theta_0) \). By the definition of the approximating cone, there exists a sequence \( \tilde{\theta} \in C\Omega_1(\theta_0) \) such that \( |\tilde{\theta} - \theta| = o(|\theta - \theta_0|) = o(n^{-1/2}) \). Let \( \tilde{h} = n^{1/2}(\tilde{\theta} - \theta_0) \). We have that \( \tilde{h} \) belongs to the cone \( C\Omega_1(\theta_0) \) and \( |\tilde{h} - h| = o(1) \). Then

\[ (I_{11}^{-1} U - h)I_{11}(I_{11}^{-1} U - h) = (I_{11}^{-1} U - \tilde{h} + \tilde{h} - h)I_{11}(I_{11}^{-1} U - \tilde{h} + \tilde{h} - h) \]

\[ \geq (I_{11}^{-1} U - \tilde{h})I_{11}(I_{11}^{-1} U - \tilde{h}) - O_p(|\tilde{h} - h|) - O_p(|\tilde{h} - h|^2) \]

\[ = (I_{11}^{-1} U - \tilde{h})I_{11}(I_{11}^{-1} U - \tilde{h}) + o_p(1). \]
Together these results imply that

\[
\inf_{h \in \Omega_n} (I_{11}^{-1}U - h)I_{11}(I_{11}^{-1}U - h) = \inf_{h \in \Omega_n} (I_{11}^{-1}U - h)I_{11}(I_{11}^{-1}U - h) + o_p(1).
\]

This completes the proof of (1). When \( \theta_0 \) is an interior point, (1) reduces to \( T(Z) = Z^T I_1 Z \) upon taking \( h = Z \), which is a weighted sum of \( d \) independent \( \chi^2_1 \) variables with the weights being the eigenvalues of \( I_{11}^{-1}I_{11}^{-1} \) by Theorem 4.4.4 of Graybill (1976). When \( \theta_0 \) is a boundary point of \( \Omega \), (1) has the same form as equation (2) in Chen & Liang (2010). Thus, the distribution of \( T(Z) \) follows from Lemma 2 of Chen & Liang (2010).

REFERENCES

Banerjee, M. (2005). Likelihood ratio tests under local alternatives in regular semiparametric models. Statist. Sinica 15, 635–44.

Bickel, P. J., Klaassen, C., Ritov, Y. & Wellner, J. (1993). Efficient and Adaptive Inference in Semiparametric Models. Baltimore, Maryland: Johns Hopkins University Press.

Chen, Y. & Liang, K.-Y. (2010). On the asymptotic behaviour of the pseudolikelihood ratio test statistic with boundary problems. Biometrika 97, 603–20.

Cheng, G. (2009). Semiparametric additive isotonic regression. J. Statist. Plan. Infer. 139, 1980–91.

Cheng, G. (2013). How many iterations are sufficient for efficient semiparametric estimation? Scand. J. Statist. 40, 592–618.

Chen, Y. & Liang, K.-Y. (2010). How many iterations are sufficient for efficient semiparametric estimation? Biometrika 97, 603–20.

Chen, G. (2009). Semiparametric additive isotonic regression. J. Statist. Plan. Infer. 139, 1980–91.

Cheng, G. (2013). How many iterations are sufficient for efficient semiparametric estimation? Scand. J. Statist. 40, 592–618.

Cheng, G. & Huang, J. Z. (2010). Bootstrap consistency for general semiparametric M-estimation. Scand. J. Statist. 38, 2884–915.

Chernoff, H. (1954). On the distribution of the likelihood ratio. Ann. Math. Statist. 25, 573–8.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65, 141–51.

Dixon, J. R., Kosorok, M. R. & Lee, B. L. (2005). Functional inference in semiparametric models using the piggyback bootstrap. Ann. Inst. Statist. Math. 57, 255–77.

Gong, G. & Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: Theory and applications. Ann. Statist. 9, 861–9.

Graybill, F. (1976). Theory and Application of the Linear Model. North Scituate, Massachusetts: Duxbury Press.

Gu, M. & Zhang, C.-H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. Ann. Statist. 21, 611–24.
Pseudolikelihood inference for semiparametric models

HOFERT, M. & MÄCHLER, M. (2011). Nested Archimedean copulas meet R: The nacopula package. *J. Statist. Software* **39**, 1–20.

HUANG, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.* **24**, 540–68.

KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer.

KUDO, A. (1963). A multivariate analogue of the one-sided test. *Biometrika* **50**, 403–18.

LIANG, K.-Y. & SELF, S. G. (1996). On the asymptotic behaviour of the pseudolikelihood ratio test statistic. *J. R. Statist. Soc. B* **58**, 785–96.

MURPHY, S. A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23**, 182–98.

MURPHY, S. A. & VAN DER VAART, A. W. (1997). Semiparametric likelihood ratio inference. *Ann. Statist.* **25**, 1471–509.

KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer.

ROBINS, J. M., ROTNITZKY, A. & ZHAO, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Am. Statist. Assoc.* **89**, 846–66.

SCHARFSTEIN, D. O., ROTNITZKY, A. & ROBINS, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J. Am. Statist. Assoc.* **94**, 1096–120.

SELF, S. G. & LIANG, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Statist. Assoc.* **82**, 605–10.

SHAPIRO, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika* **72**, 133–44.

SIIH, J. H. & LOUIS, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics* **51**, 1384–99.

SONG, R., KOSOROK, M. R. & FINE, J. P. (2009). On asymptotically optimal tests under loss of identifiability in semiparametric models. *Ann. Statist.* **37**, 2409–44.

VAN DER VAART, A. W. (2000). *Asymptotic Statistics*. New York: Cambridge University Press.

WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā A* **26**, 359–72.

WONG, W. H. & SEVERINI, T. A. (1991). On maximum likelihood estimation in infinite dimensional parameter spaces. *Ann. Statist.* **19**, 603–32.

YAN, J. (2007). Enjoy the joy of copulas: With a package copula. *J. Statist. Software* **21**, 1–21.

[Received on 6 May 2014. Editorial decision on 15 October 2016]