

A conditional composite likelihood ratio test with boundary constraints

By YONG CHEN, JING HUANG

*Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania,
Philadelphia, Pennsylvania 19104, U.S.A.*
ychen123@upenn.edu jing14@upenn.edu

YANG NING

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

KUNG-YEE LIANG

Department of Life Sciences, National Yang-Ming University, Li-Nong Street, Taipei City 112, Taiwan
kyluang@nhri.org.tw

AND BRUCE G. LINDSAY

SUMMARY

Composite likelihood has been widely used in applications. The asymptotic distribution of the composite likelihood ratio statistic at the boundary of the parameter space is a complicated mixture of weighted χ^2 distributions. In this paper we propose a conditional test with data-dependent degrees of freedom. We consider a modification of the composite likelihood which satisfies the second-order Bartlett identity. We show that the modified composite likelihood ratio statistic given the number of estimated parameters lying on the boundary converges to a simple χ^2 distribution. This conditional testing procedure is validated through simulation studies.

Some key words: Boundary problem; Composite likelihood; Likelihood ratio test; Nonstandard inference.

1. INTRODUCTION

Composite likelihood (Besag, 1974; Lindsay, 1988) is an inference function constructed as the product of a set of conditional and/or marginal density functions. Composite likelihoods have been widely used in longitudinal studies, spatial modelling, analysis of panel data or missing data, and other areas. When a working independence assumption is adopted, the composite likelihood is sometimes called the independence likelihood (Chandler & Bate, 2007). Varin et al. (2011) reviewed composite likelihood methods.

Although Wald-based inference using composite likelihoods is common, likelihood ratio-based inference often has better finite-sample performance. The composite likelihood ratio statistic is generally distributed as a linear combination of independent χ^2_1 distributions (Molenberghs & Verbeke, 2005). This is mainly due to the failure of the second-order Bartlett identity, which refers to the equality between the expected negative Hessian of the loglikelihood and the covariance of the score function. To deal with such complications, several adjustments have been proposed (Chandler & Bate, 2007; Pace et al., 2011), and the adjusted composite likelihood ratio statistics converge weakly to a simple χ^2 distribution.

A regularity condition underlying these adjustments is that the parameter θ_0 under the null hypothesis $H_0 : \theta = \theta_0$ is interior to its parameter space, so that the composite score function is zero at the maximum composite likelihood estimate. However, when the parameter θ_0 lies on the boundary of the parameter space, which commonly occurs in variance component models, this first-order condition is usually not satisfied. Consequently, existing adjustments for composite likelihoods may not yield an asymptotic χ^2 distribution. Another challenge for boundary problems is that the asymptotic distribution is often a mixture of χ^2 distributions (Self & Liang, 1987; Chen & Liang, 2010; Chen et al., 2017), and calculation of the mixing proportions relies on a partition of the parameter space. Such a partition often depends on the geometry of the tangent cone at the null hypothesis as well as on the decomposition of the Fisher information, and must be worked out case by case; this limits the use of these asymptotic results.

The primary aim of this paper is to develop a modified composite likelihood ratio statistic with a simple limiting distribution for hypothesis testing when the parameter of interest lies on the boundary of the parameter space. Specifically, to recover the second-order Bartlett identity, we propose a novel modification of the composite likelihood, whereby a new approximation is obtained via a quadratic function of a linear combination of the composite score function and the maximum composite likelihood estimator. To avoid the calculation of mixing proportions, we adopt a conditional testing procedure recently proposed by Susko (2013), which originated from the ideas of Bartholomew (1961). Susko showed that the standard likelihood ratio statistics given the number of parameters lying on the boundary converges weakly to a simple χ^2 distribution with data-dependent degrees of freedom. A crucial assumption made in Bartholomew (1961) and Susko (2013) is the second-order Bartlett identity, which does not hold for composite likelihoods. The goals of this work are two-fold: we extend Chandler & Bate's (2007) adjustment to boundary problems, and we generalize Susko's (2013) result to more general likelihoods for which the second-order Bartlett identity does not hold. We give a general theorem on the asymptotic distribution of the modified composite likelihood ratio statistic, and we show empirically that the modified composite likelihood ratio statistic performs better than the naive test that ignores boundary problems.

2. MODIFIED COMPOSITE LIKELIHOOD UNDER BOUNDARY CONSTRAINTS

Let $g(x; \theta)$ be the probability density function of a multi-dimensional random vector X , indexed by a p -dimensional parameter $\theta = (\theta_1, \dots, \theta_p)^T$, where θ belongs to the parameter space Ω . We assume that distinct values of θ correspond to distinct probability distributions. Let $\{\mathcal{A}_1, \dots, \mathcal{A}_K\}$ denote a set of marginal or conditional events associated with loglikelihoods $\ell_k\{\theta; \mathcal{A}_k(x)\} = \log \int_{x \in \mathcal{A}_k} g(x; \theta) dx$, where $k = 1, \dots, K$ with K being the total number of events. Suppose that N independent random variables X_1, \dots, X_N are observed from the model $g(x; \theta)$. Following Lindsay (1988), a composite loglikelihood can be constructed as

$$\ell_c(\theta) = \sum_{i=1}^N \sum_{k=1}^K \omega_{ik} \ell_k\{\theta; \mathcal{A}_k(x_i)\},$$

where ω_{ik} is a nonnegative deterministic weight associated with the loglikelihood component $\ell_k\{\theta; \mathcal{A}_k(x_i)\}$. Let $\hat{\theta}_c = \arg \max_{\theta \in \Omega} \ell_c(\theta)$ be the maximum composite likelihood estimator, and define the composite score function, sensitivity matrix and variability matrix as, respectively,

$$U_c(\theta) = \frac{\partial \ell_c(\theta)}{\partial \theta}, \quad H = \lim_{N \rightarrow \infty} -\frac{1}{N} E \left\{ \frac{\partial^2 \ell_c(\theta)}{\partial \theta^T \partial \theta} \right\}, \quad V = \lim_{N \rightarrow \infty} \frac{1}{N} E \left[\left\{ \frac{\partial \ell_c(\theta)}{\partial \theta} \right\} \left\{ \frac{\partial \ell_c(\theta)}{\partial \theta} \right\}^T \right].$$

The corresponding estimators of H and V are denoted by \hat{H} and \hat{V} , evaluated at $\hat{\theta}_c$. The second-order Bartlett identity generally does not hold for $\ell_c(\theta)$, since $H \neq V$. To circumvent this problem,

Chandler & Bate (2007) proposed an adjusted composite likelihood through a vertical scaling:

$$\ell_A(\theta) = \ell_c(\hat{\theta}_c) + \left\{ (\theta - \hat{\theta}_c)^T \hat{H}_A (\theta - \hat{\theta}_c) \right\} \frac{\ell_c(\theta) - \ell_c(\hat{\theta}_c)}{(\theta - \hat{\theta}_c)^T \hat{H} (\theta - \hat{\theta}_c)}, \tag{1}$$

where \hat{H}_A is the inverse of the robust variance estimator $\hat{H}\hat{V}^{-1}\hat{H}$ and $(\theta - \hat{\theta})^T \hat{H}_A (\theta - \hat{\theta}) / (\theta - \hat{\theta})^T \hat{H} (\theta - \hat{\theta})$ is the scaling factor. Pace et al. (2011) proposed another adjustment based on the composite score function.

However, when the parameter θ_0 of the null hypothesis lies on the boundary of its parameter space, the two adjustments above are not sufficient to enforce the second-order Bartlett identity at θ_0 , because both are based on approximation of $\ell_c(\hat{\theta}_c)$ around θ_0 by a quadratic form of $(\hat{\theta}_c - \theta_0)$ (Chandler & Bate, 2007) or a quadratic form of $U_c(\theta_0)$ (Pace et al., 2011), which relies on the asymptotic equivalence of $N^{1/2}(\hat{\theta}_c - \theta_0)$ and $N^{-1/2}\hat{H}^{-1}U_c(\theta_0)$. Such an equivalence is true only under the first-order condition $U_c(\hat{\theta}_c) = 0$. Under boundary constraints, $\hat{\theta}_c$ may not solve this composite score equation, so the asymptotic equivalence does not hold. We therefore adapt the quadratic approximation in Lemma 1 of Self & Liang (1987) to deal with composite likelihood, and obtain the following approximation around $\hat{\theta}_c$:

$$2 \left\{ \ell_c(\theta) - \ell_c(\hat{\theta}_c) \right\} = -T(\theta)^T \hat{H} T(\theta) + N^{-1} U_c(\hat{\theta}_c)^T \hat{H}^{-1} U_c(\hat{\theta}_c) + O_p \left(N \|\theta - \hat{\theta}_c\|^3 \right), \tag{2}$$

where $T(\theta) = N^{-1/2} \hat{H}^{-1} U_c(\hat{\theta}_c) - N^{1/2} (\theta - \hat{\theta}_c)$. Under boundary constraints, the composite likelihood $\ell_c(\theta)$ is approximated by a quadratic form of a linear combination of $U_c(\hat{\theta}_c)$ and $(\theta - \hat{\theta}_c)$.

In a similar spirit to Chandler & Bate (2007) and Pace et al. (2011), we propose the following modified composite likelihood under boundary constraints:

$$\ell_M(\theta) = \ell_c(\hat{\theta}_c) - \left\{ T(\theta)^T \hat{H}_A T(\theta) \right\} \phi(\theta), \tag{3}$$

where $T(\theta) = N^{-1/2} \hat{H}^{-1} U_c(\hat{\theta}_c) - N^{1/2} (\theta - \hat{\theta}_c)$ and

$$\phi(\theta) = \frac{\ell_c(\theta) - \ell_c(\hat{\theta}_c)}{-T(\theta)^T \hat{H} T(\theta) + N^{-1} U_c(\hat{\theta}_c)^T \hat{H}^{-1} U_c(\hat{\theta}_c)}.$$

In (3), the modified composite likelihood $\ell_M(\theta)$ is approximated by a quadratic form with a finite-sample adjustment $\phi(\theta)$. The quadratic term vertically scales the likelihood to enforce the second-order Bartlett identity. The proposed modification uses a second-order expansion based on both $U_c(\hat{\theta}_c)$ and $(\theta - \hat{\theta}_c)$ to deal with boundary constraints. When there is no boundary constraint, the proposed modified composite likelihood reduces to Chandler & Bate's (2007) adjustment in (1) upon noting that $U_c(\hat{\theta}_c) = 0$.

3. THEORETICAL RESULTS

We decompose the p -dimensional parameter θ as $\theta^T = (\zeta^T, \eta^T)$, where the null hypothesis $H_0 : \zeta = 0$ and $\eta = \eta_0$ constrains the s -dimensional parameter ζ to lie on the boundary of the parameter space and sets the remaining $(p - s)$ -dimensional parameter η to a value in the interior of the parameter space. The composite likelihood ratio statistic based on the proposed modification is

$$W = -2 \left\{ \ell_M(0, \eta_0) - \ell_M(\hat{\theta}_M) \right\}.$$

Using the quadratic expansion for boundary parameters in (2) and the projections of normal random vectors onto cones in Shapiro (1985), we obtain our main results.

THEOREM 1. *Under the regularity conditions R1–R5 in the Supplementary Material, the following properties hold.*

- (i) The second-order Bartlett identity holds asymptotically for $\ell_M(\theta)$ at θ_0 , i.e., $N^{-1} \text{var}\{\partial \ell_M(\theta_0)/\partial \theta\} = -N^{-1} E\{\partial^2 \ell_M(\theta_0)/\partial \theta^T \partial \theta\} + O_p(N^{-1/2})$.
- (ii) With probability tending to 1, as $N \rightarrow \infty$, there exists a sequence of points $\hat{\theta}_M$ in the parameter space Ω where local maxima of $\ell_M(\theta)$ occur and which converges to θ_0 in probability. Moreover, $N^{1/2}(\hat{\theta}_M - \theta_0) = O_p(1)$.
- (iii) Let V denote the number of null hypothesis parameters that are estimated to be in the interior of the parameter space and let v denote its observed value. Then

$$\text{pr}(W \leq w \mid V = v) \rightarrow \text{pr}(\chi_v^2 \leq w)$$

as $N \rightarrow \infty$, where χ_v^2 is a chi-squared variable with v degrees of freedom.

This theorem shows that, given $V = v$, the modified composite likelihood ratio statistic converges weakly to χ_v^2 , which extends the results in Susko (2013). Without the modification of the composite likelihood, one could conduct a standard composite likelihood ratio test. The calculation of the asymptotic distribution of the test statistic under boundary constraints follows an argument similar to that in Chen & Liang (2010), where pseudolikelihood (Gong & Samaniego, 1981) is considered and the second-order Bartlett identity also does not hold. However, the calculation involves projection onto the tangent cone at the null hypothesis, as well as decomposition of the Fisher information, and must be worked out case by case. Another alternative is to calculate the composite likelihood ratio statistic and compare it with the χ_p^2 distribution. This is a naive implementation of the composite likelihood ratio test, since the boundary constraints are ignored and the second-order Bartlett identity is taken to be valid even though it is not. Unlike in standard likelihood inference, the naive test using composite likelihood is not necessarily conservative, because the distribution of the composite likelihood ratio statistic converges to a mixture of distributions, some of which may have heavier tails than χ^2 (Chen & Liang, 2010). The proposed modified composite likelihood ratio test is a good compromise, and its calculation is straightforward and easy to implement.

4. SIMULATION

We conducted simulations to validate the theoretical results empirically. Our illustrative example is a hypothesis testing problem in stratified case-control studies (Liang, 1987).

In a stratified case-control study, let x_{i1}, \dots, x_{ik_i} denote the $p \times 1$ vectors of potential risk factors for k_i cases, and let $x_{ik_i+1}, \dots, x_{iK_i}$ denote the potential risk factors of $K_i - k_i$ controls in the i th stratum ($i = 1, \dots, N$). A logistic regression model allowing for stratum-specific effects is

$$\text{logit pr}(y_{ij} = 1 \mid x_{ij}) = \alpha_i + \beta^T x_{ij} \quad (i = 1, \dots, N), \tag{4}$$

where the coefficients β quantify the effects of the risk factors x_{ij} on the disease status y_{ij} .

To draw valid inference on β , Liang (1987) proposed a composite likelihood method where the nuisance parameters α_i are eliminated by conditioning. For the (j, l) th case-control pair of subjects in the i th stratum ($j = 1, \dots, k_i; l = k_i + 1, \dots, K_i$), the conditional probability that x_{ij} is from the case given that x_{ij} and x_{il} are observed is

$$\text{pr}(y_{ij} = 1, y_{il} = 0 \mid y_{ij} + y_{il} = 1, x_{ij}, x_{il}; \alpha_i, \beta) = \frac{\exp(\beta^T x_{ij})}{\exp(\beta^T x_{ij}) + \exp(\beta^T x_{il})}.$$

Hence, a composite loglikelihood combining all possible pairs from N strata with weights w_{ijl} is

$$\ell_c(\beta) = \sum_{i=1}^N \sum_{j=1}^{k_i} \sum_{l=k_i+1}^{K_i} w_{ijl} \log \left\{ \frac{\exp(\beta^T x_{ij})}{\exp(\beta^T x_{ij}) + \exp(\beta^T x_{il})} \right\}.$$

Table 1. Type I errors ($\times 100$) of the modified composite likelihood ratio test and the naive composite likelihood ratio test, for $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ at nominal levels 0.05 and 0.1, where $\text{logit pr}(y_{ij} = 1 | x_{ij}) = \alpha_i$ and $\alpha_i \sim \text{Un}[-1, 1]$

	Level (%)	Empirical/Theoretical	Number of strata				
			25	50	100	200	400
mCLRT	5	Empirical	5.0	5.3	5.1	6.1	4.9
		Theoretical	4.3	4.4	4.4	4.4	4.4
	10	Empirical	8.3	10.1	10.1	10.7	9.5
		Theoretical	8.7	8.8	8.8	8.8	8.8
Naive	5	Empirical	0.2	0.0	0.1	0.0	0.0
		Theoretical	5.0	5.0	5.0	5.0	5.0
	10	Empirical	0.2	0.3	0.2	0.4	0.1
		Theoretical	10.0	10.0	10.0	10.0	10.0

mCLRT, the modified composite likelihood ratio test; Naive, the naive composite likelihood ratio test; Empirical, the empirical Type I error; Theoretical, the theoretical false positive probability.

Without loss of generality we set $w_{ijl} = K_i^{-1}$, so the maximum composite likelihood estimator reduces to the Mantel–Haenszel estimator when a binary covariate is considered (Liang, 1987). Suppose that the covariates are three-dimensional and are known to be positively associated with the occurrence of disease, i.e., the null hypothesis is $H_0 : \beta = (0, 0, 0)$ and the alternative is $H_a : \text{any of } \beta_1, \beta_2, \beta_3 > 0$. For this problem, the standard composite likelihood ratio statistic does not converge to χ_3^2 because of the boundary constraint. Setting $C_{\Omega_0} = \{0\} \times \{0\} \times \{0\}$ and $C_{\Omega} = [0, \infty) \times [0, \infty) \times [0, \infty)$, the asymptotic distribution of the composite likelihood ratio statistic can be derived following Chen & Liang (2010), given that the test statistic is asymptotically equivalent to the difference between two quadratic forms, with

$$Z^T H_0 Z - \inf_{\beta \in C_{\Omega}} (Z - \beta)^T H_0 (Z - \beta), \quad Z \sim N(0, H_0^{-1} V_0 H_0^{-1}).$$

As β is three-dimensional, the calculations of each individual distribution and of the mixing proportions in the limiting mixture are rather complicated. In this situation, a naive comparison of the composite likelihood ratio statistic with a χ_3^2 distribution is a simple procedure, so we focus on comparing the modified composite likelihood ratio test and the naive tests.

We simulate data from several different scenarios and compare the Type I error and power of the modified composite likelihood ratio test with those of the naive composite likelihood ratio test. The naive test uses the $(1 - \alpha)$ th quantile of χ_3^2 as the critical value, where α is the nominal level, and the modified composite likelihood ratio test uses the $(1 - \alpha)$ th quantile of χ_v^2 , where v is the number of β coefficients that were estimated as positive. For all the scenarios, the number of cases and the number of controls for each stratum are fixed at five, whereas the number of strata varies from 25 to 400. For each stratum, we generate a stratum-specific intercept α_i from a uniform distribution on $[-1, 1]$, and generate covariates x_1, x_2 and x_3 from a standard normal distribution. We then calculate the individual probability of having disease using (4) and generate the binary disease status from the corresponding Bernoulli distribution. Finally, we randomly sample five cases and five controls from each stratum. Box-constrained optimization of the likelihoods is implemented through the R function `optim` (R Development Core Team, 2018) with the L-BFGS-B method. We use 10^{-6} as the threshold to determine whether the estimated parameter is on the boundary. Sensitivity analyses yield similar results when the threshold is varied from 10^{-3} to 10^{-8} .

Table 1 summarizes the Type I errors based on 5000 simulations. The naive composite likelihood ratio test has empirical Type I errors that are much smaller than the nominal levels, but the modified composite likelihood ratio test performs reasonably well. Since the modified likelihood ratio test cannot reject the null hypothesis when all the parameter estimates are on the boundary, the theoretical false positive probabilities for the proposed test corresponding to a nominal level of α can only be attained at $\alpha \times \{1 - \text{pr}(\text{all } \hat{\beta} = 0)\}$; these probabilities are listed in the second and fourth rows in the mCLRT part of Table 1.

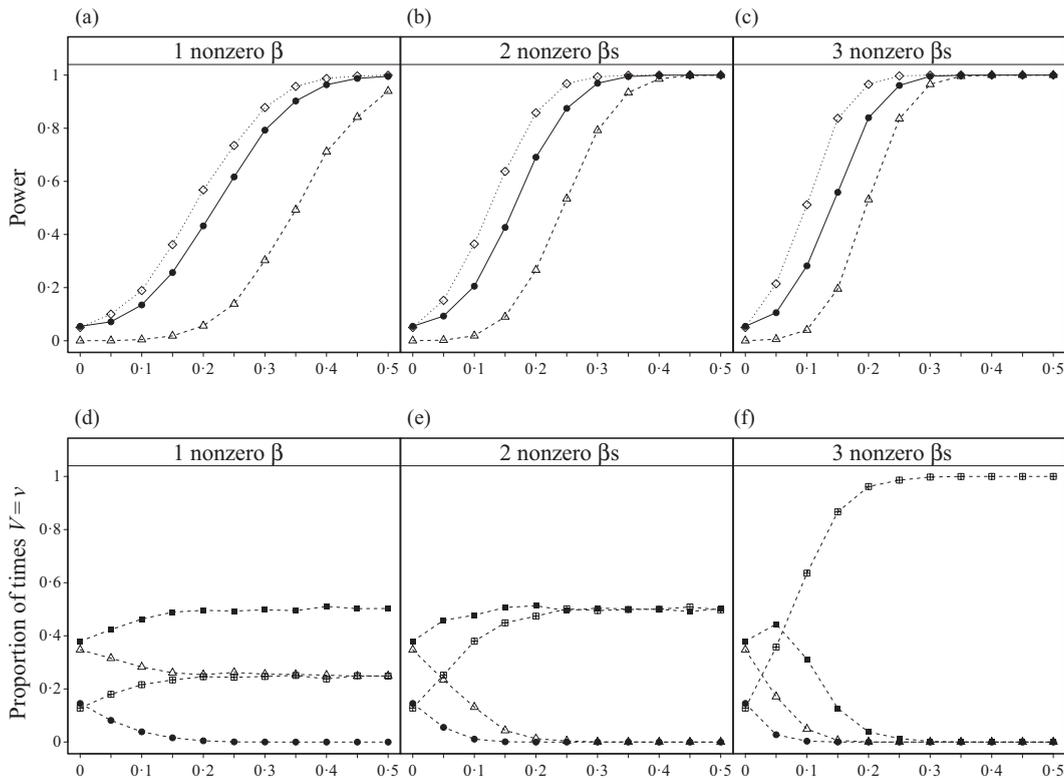


Fig. 1. Power comparisons for the matched case-control study example when one, two or three testing parameters are positive in the data-generating model: (a)–(c) power of the modified composite likelihood ratio test (\bullet), the naive test (Δ), and the composite likelihood ratio test (\diamond) as the nonzero effect size increases from 0 to 0.5, with the number of strata equal to 50; (d)–(f) the proportion of times that β coefficients were estimated as positive when $V = 0$ (\bullet), 1 (Δ), 2 (\blacksquare) or 3 (\boxplus) as the nonzero effect size increases from 0 to 0.5, with the number of strata equal to 50.

To compare the power, we let one, two or three β coefficients vary from 0.0 to 0.5 with 50 strata. For the nonzero β coefficients, we assume equal effect sizes. Panels (a)–(c) of Fig. 1 show the power curves at a 5% nominal level based on 5000 simulations; panels (d)–(f) show the frequencies of $V = 0, 1, 2$ and 3 as the nonzero effect size increases from 0.0 to 0.5. The proposed test always has larger power than the naive test. The gain in power can be further explained by the proportion of times that β is estimated as positive. Specifically, when all three β coefficient estimates are far from the boundary in the data-generating model, most of the time we obtain three positive β estimates. Thus, both the naive test and the proposed test refer to χ^2_3 and perform more similarly, as illustrated in Fig. 1(c) and (f). On the other hand, when there are only one or two nonzero or three nonzero but small β coefficient estimates, the proposed test frequently uses one or two degrees of freedom, even as the additional β values get larger; see Fig. 1. There is a substantial gain in power in these situations.

We also investigate the potential power loss due to the conditioning procedure and the modification. The critical value of the composite likelihood ratio test is obtained by calculating the 95% quantile of 5000 test statistics based on data generated under the null hypothesis. As indicated by Fig. 1(a)–(c), there is a moderate loss of power. However, theoretical calculation of the critical value of the composite likelihood ratio test can be tedious and must be done case by case. Therefore, the modified composite likelihood ratio test serves as a useful alternative at the cost of moderate power loss.

5. DISCUSSION

The proposed modification also applies to the composite likelihood when nuisance parameters are present. Assume that $\gamma^T = (\theta^T, \lambda^T)$, where λ is a finite-dimensional nuisance parameter and $\theta^T = (\zeta^T, \eta^T)$.

Similarly, we test the null hypothesis $H_0 : \zeta = 0$ and $\eta = \eta_0$, where 0 is on the boundary of the parameter space for ζ and η_0 is in the interior of the parameter space. Let $\ell_{pc}(\theta) = \ell_c(\theta, \hat{\lambda}_\theta)$ be the profile composite likelihood for θ , where $\hat{\lambda}_\theta = \arg \max_\lambda \ell_c(\theta, \lambda)$. Let

$$U_p(\theta) = \frac{\partial \ell_{pc}(\theta)}{\partial \theta}, \quad H_p = \lim_{N \rightarrow \infty} -\frac{1}{N} E \left\{ \frac{\partial^2 \ell_{pc}(\theta)}{\partial \theta^T \partial \theta} \right\},$$

$$V_p = \lim_{N \rightarrow \infty} \frac{1}{N} E \left[\left\{ \frac{\partial \ell_{pc}(\theta)}{\partial \theta} \right\} \left\{ \frac{\partial \ell_{pc}(\theta)}{\partial \theta} \right\}^T \right]$$

be, respectively, the profile composite score function, sensitivity matrix and variability matrix for θ , and let \hat{H}_p and \hat{V}_p denote the corresponding estimates; also define $\hat{H}_{pA} = \hat{H}_p \hat{V}_p^{-1} \hat{H}_p$. The modified profile composite likelihood is $\ell_{MP}(\theta) = \ell_{pc}(\hat{\theta}_c) - \{T_p(\theta)^T \hat{H}_{pA} T_p(\theta)\} \phi_p(\theta)$ where $T_p(\theta) = N^{-1/2} \hat{H}_p^{-1} U_{pc}(\hat{\theta}_c) - N^{1/2}(\theta - \hat{\theta}_c)$ and

$$\phi_p(\theta) = \frac{\ell_{pc}(\theta) - \ell_{pc}(\hat{\theta}_c)}{-T_p(\theta)^T \hat{H}_p T_p(\theta) + N^{-1} U_{pc}(\hat{\theta}_c)^T \hat{H}_p^{-1} U_{pc}(\hat{\theta}_c)}.$$

We define the modified profile composite likelihood ratio statistic to be $W_p = 2\{\ell_{MP}(\hat{\theta}_M) - \ell_{MP}(0, \eta_0)\}$, where $\hat{\theta}_M = \arg \max_\theta \ell_{MP}(\theta)$. Here \hat{H}_p and \hat{V}_p can be calculated empirically; see the Supplementary Material.

When the parameter is on the boundary of the parameter space, Andrews (2000) established the inconsistency of the standard nonparametric and parametric bootstrap methods and proposed subsampling and m -out-of- n bootstrap methods for obtaining consistent estimators of the limiting distribution. However, both these methods require additional tuning parameters. Our proposed method is computationally simpler and free of unknown tuning parameters.

ACKNOWLEDGEMENT

The first four authors are truly indebted to the fifth author, Professor Bruce G. Lindsay of Pennsylvania State University, for his inspiration and encouragement; he will be remembered as a great scholar. The authors thank the referee, associate editor and editor for comments that substantially improved this work. This research was supported in part by the National Institutes of Health, U.S.A.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes regularity conditions, the proof of the asymptotic expansion in equation (2), the proof of Theorem 1, and a more general result involving nuisance parameters.

REFERENCES

- ANDREWS, D. W. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* **68**, 399–405.
- BARTHOLOMEW, D. J. (1961). A test of homogeneity of means under restricted alternatives. *J. R. Statist. Soc. B* **23**, 239–81.
- BESAG, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. R. Statist. Soc. B* **36**, 192–236.
- CHANDLER, R. E. & BATE, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika* **94**, 167–83.
- CHEN, Y. & LIANG, K.-Y. (2010). On the asymptotic behaviour of the pseudolikelihood ratio test statistic with boundary problems. *Biometrika* **97**, 603–20.
- CHEN, Y., NING, J., NING, Y., LIANG, K.-Y. & BANDEEN-ROCHE, K. (2017). On pseudolikelihood inference for semiparametric models with boundary problems. *Biometrika* **104**, 165–79.

- GONG, G. & SAMANIEGO, F. J. (1981). Pseudo maximum likelihood estimation: Theory and applications. *Ann. Statist.* **9**, 861–9.
- LIANG, K.-Y. (1987). Extended Mantel-Haenszel estimating procedure for multivariate logistic regression models. *Biometrics* **43**, 289–99.
- LINDSAY, B. G. (1988). Composite likelihood methods. *Contemp. Math.* **80**, 221–39.
- MOLENBERGHS, G. & VERBEKE, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.
- PACE, L., SALVAN, A. & SARTORI, N. (2011). Adjusting composite likelihood ratio statistics. *Statist. Sinica* **21**, 129–48.
- R DEVELOPMENT CORE TEAM (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org>.
- 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.
- SUSKO, E. (2013). Likelihood ratio tests with boundary constraints using data-dependent degrees of freedom. *Biometrika* **100**, 1019–23.
- VARIN, C., REID, N. & FIRTH, D. (2011). An overview of composite likelihood methods. *Statist. Sinica* **21**, 5–42.

[Received on 12 May 2015. Editorial decision on 3 September 2017]