Biometrika (2015), **99**, 1, *pp*. 1–8

Printed in Great Britain

Advance Access publication on 31 July 2015

A conditional composite likelihood ratio test with boundary constraints

BY YONG CHEN

Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, U.S.A.

ychen123@upenn.edu

JING HUANG

Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, U.S.A.

jing14@upenn.edu

YANG NING

Department of Statistical Science, Cornell University, Ithaca, New York 14853, U.S.A.

yn265@cornell.edu

KUNG-YEE LIANG

Department of Life Sciences, National Yang-Ming University, Taipei City 112, Taiwan

kyliang@nhri.org.tw

AND BRUCE G. LINDSAY

Department of Statistics, The Pennsylvania State University, State College, Pennsylvania 16802, U.S.A.

SUMMARY

Composite likelihood has been widely used in applications. 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 composite likelihood, which achieves 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; non-standard condition.

1. INTRODUCTION

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

Although Wald-based inference has been established, likelihood based inference is sometimes preferred for better finite sample performance. The composite likelihood ratio statistic is generally distributed as a linear combination of independent χ_1^2 distributions (Molenberghs & Verbeke, 2005). This is mainly due to the failure of the second-order Bartlett identity. Recall that the second-order Bartlett identity refers to the equality between the expected value of negative Hessian of log likelihood and the covariance of score function. To deal with such complication, several adjustments have been proposed (Chandler & Bate, 2007; Pace et al., 2011), and the adjusted composite likelihood ratio statistic converges weakly to a simple χ^2 distribution. For example, Chandler & Bate (2007) proposed a vertical scaling approach to stretch the composite likelihood to restore the second-order Bartlett identity. Pace et al. (2011) proposed another modification based on the composite score function.

A regularity condition underlying these adjustments of composite likelihood is that the parameter θ_0 under the null hypothesis $H_0 : \theta = \theta_0$ is interior to its parameter space, so 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 is common for variance component models, this first-order condition is not satisfied. Consequently, the 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), where 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 the decomposition of Fisher information, and has to be worked out case by case. This limits the use of these asymptotic results.

The primary purpose of this paper is to propose a modified composite likelihood ratio statistic with a simple limiting distribution for hypothesis testing when the parameter of interest lies on the boundary of its space. Specifically, to recover the second-order Bartlett identity, we propose a novel modification of composite likelihood, adapting a new approximation through a quadratic function of a linear combination of the composite score function and maximum composite likelihood estimator. To avoid the calculation of mixing proportions, we adopt a conditional testing procedure recently proposed by Susko (2013), originated from Bartholomew (1961). Susko showed that the standard likelihood ratio statitics 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 the argument of Bartholomew (1961) and Susko (2013) is the second-order Bartlett identity, which does not hold for composite likelihoods. The purposes of this work are twofold. First, we extend Chandler & Bate (2007)'s adjustment to boundary problems. Second, we generalize Susko's result to more generally defined likelihoods where the second-order Bartlett identity does not hold. We provide a general theorem on the asymptotic distribution of the modified composite likelihood ratio statistic, and 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 multidimensional random vector X, indexed by a p-dimensional parameter $\theta = (\theta_1, \ldots, \theta_p)^{\mathrm{T}}$, where θ belongs to the parameter space Ω. We assume that distinct values of θ correspond to distinct probability distributions. Let $\{\mathscr{A}_1, \ldots, \mathscr{A}_K\}$ denote a set of marginal or conditional events associated with log likelihoods

$\ell_k\{\theta; \mathscr{A}_k(x)\} = \log \int_{x \in \mathscr{A}_k} g(x;\theta)dx$, where $k = 1, \ldots, K$ and K is the total number of events. Suppose N independent random variables $X_1, \ldots, X_N$ are observed from the model $g(x;\theta)$. Following Lindsay (1988), a composite log likelihood can be constructed as

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

where ω_{ik} is a nonnegative deterministic weight associated with the log likelihood $\ell_k\{\theta; \mathscr{A}_k(x_i)\}$. Let $\hat{\theta}_c = \arg\max_{\theta \in \Omega} \ell_c(\theta)$ be the maximum composite likelihood estimator, and

$$U_c(\theta) = \frac{\partial \ell_c(\theta)}{\partial \theta}, \;\; H = \lim_{N\to\infty} -\frac{1}{N} E\left\{\frac{\partial^2 \ell_c(\theta)}{\partial \theta^{\mathrm{T}} \partial \theta}\right\}, \;\; V = \lim_{N\to\infty} \frac{1}{N} E\left\{\frac{\partial \ell_c(\theta)}{\partial \theta}\right\}\left\{\frac{\partial \ell_c(\theta)}{\partial \theta}\right\}^{\mathrm{T}}$$

be the composite score function, sensitivity matrix and variability matrix, respectively. 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 as

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

where $\hat{H}_A$ is the inverse of robust variance estimator $\hat{H}\hat{V}^{-1}\hat{H}$, and $(\theta - \hat{\theta})^{\mathrm{T}} \hat{H}_A (\theta - \hat{\theta})/(\theta - \hat{\theta})^{\mathrm{T}} \hat{H} (\theta - \hat{\theta})$ 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 above adjustments 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, i.e., $U_c(\hat{\theta}_c) = 0$. 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 composite likelihood, and obtain the following approximation around $\hat{\theta}_c$,

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

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 as in Chandler & Bate (2007) and Pace et al. (2011), we propose the following modified composite likelihood under boundary constraints

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

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

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

In equation (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 adjustment in equation (1) by noting $U_c(\hat{\theta}_c) = 0$.

3. THEORETICAL RESULTS

We decompose the p-dimensional parameter θ as $\theta^{\mathrm{T}} = (\zeta^{\mathrm{T}}, \eta^{\mathrm{T}})$, where the null hypothesis $H_0 : \zeta = 0$ and $\eta = \eta_0$ constrains the s-dimensional parameter ζ to 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 equation (2) and the projections of normal random vectors onto cones in Shapiro (1985), we can obtain our main results as follows.

THEOREM 1. *Under the conditions in Section 3.1 and the regularity conditions R1-R5 in Supplementary Material, we have the following results.*

1. *The second-order Bartlett identity holds asymptotically for $\ell_M(\theta)$ at θ_0, i.e. $N^{-1}var\{\partial\ell_M(\theta_0)/\partial\theta\} = -N^{-1}E\left\{\partial^2\ell_M(\theta_0)/\partial\theta^{\mathrm{T}}\partial\theta\right\} + \mathrm{O_p}(N^{-1/2})$;*
2. *With probability tending to 1, as $N \to \infty$, there exists a sequence of points in the parameter space Ω, $\hat{\theta}_M$, at which local maxima of $\ell_M(\theta)$ occur, and that converges to θ_0 in probability. Moreover, $N^{1/2}(\hat{\theta}_M - \theta_0) = \mathrm{O_p}(1)$;*
3. *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. We have*

$$pr(W \leq w \mid V = v) \to pr(\chi^2_v \leq w), \quad N \to \infty,$$

where χ^2_v 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^2_v, which extends the results in Susko (2013). Without the modification of composite likelihood, one could conduct a standard composite likelihood ratio test. The calculation of the asymptotic distribution of test statistic under boundary constraints follows an argument similar to 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 the decomposition of Fisher information, and must be worked out case by case. Another alternative is to calculate the composite likelihood ratio statistic and compare it to the χ^2_p 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 pretended to be valid. 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, whose calculation is straightforward and easy to implement.

4. SIMULATION

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

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

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

where the coefficients β quantify the effects of 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) case-control pair of subjects of the ith stratum $(j = 1, \ldots, k_i; l = k_i + 1, \ldots, K_i)$, the conditional probability that x_{ij} is from the case given that x_{ij} and x_{il} are observed is

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

Thus, a composite log likelihood combining all possible pairs from N strata with weight w_{ijl} is

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

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 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 : any of $\beta_1, \beta_2, \beta_3 > 0$. For this problem, the standard composite likelihood ratio statistic does not converge to χ^2_3 due to the boundary constraint. Setting $C_{\Omega_0} = \{0\} \times \{0\} \times \{0\}$, $C_\Omega = [0, \infty) \times [0, \infty) \times [0, \infty)$, the asymptotic distribution of the composite likelihood ratio statistic can be derived following Chen & Liang (2010), in which the test statistic is asymptotically equivalent to the difference between two quadratic forms, with

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

As β is three-dimensional, the calculations of each individual distribution and the mixing proportions in the limiting mixture are rather complicated. In this situation, a naive comparison of the composite likelihood ratio statistic to a χ^2_3 distribution is a simple procedure, so we focus on the comparison between 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 to the naive composite likelihood ratio test. The naive test uses the $(1 - \alpha)$th quantile of χ^2_3 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^2_v, where v is the number of β coefficients that were estimated as positive. For all the scenarios, the number of cases and number of controls for each stratum are fixed to be 5, 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 equation (4) and generate the binary disease status from the corresponding Bernoulli distribution. Finally, we randomly sample 5 cases and 5 controls from each stratum. Box-constrained optimization of the likelihoods is implemented through the R function optim with the L-BFGS-B method. We use 10^{-6} as the threshold to determine if the estimated parameter is on the boundary. Sensitivity analyses show similar results when the threshold is varied from 10^{-3} to 10^{-8}.

Table 1. *Type I errors of the modified composite likelihood ratio test (mCLRT) and the naive composite likelihood ratio test (Naive) for* $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ *at 0·05 and 0·1 nominal levels, where logit pr*$(y_{ij} = 1 \mid x_{ij}) = \alpha_i$, *and* $\alpha_i \sim U[-1, 1]$. *All entries are multiplied by 100.*

			Number of Strata				
	Level(%)	Empirical/Theoretical	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

Empirical, the empirical type I error; Theoretical, the theoretical false positive probability.

Table 1 summarizes the type I errors based on 5,000 simulations. The naive composite likelihood ratio test has empirical type I errors 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 nominal level of α can only be attained at $\alpha \times \{1 - \text{pr(all } \hat{\beta} = 0)\}$, which are listed in the second row in the upper panel of Table 1.

To compare the power, we let one, two, or three β coefficients vary from 0·0 to 0·5 with 50 strata. For the non-zero β coefficients, we assume equal effect sizes. The first row of Figure 1 shows the power curves at 5% nominal level based on $5,000$ simulations. The second row shows the frequencies of $V = 0, 1, 2,$ or 3 as the non-zero 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 probability 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 the panels (c) and (f) of Figure 1. On the other hand, when there are only one or two nonzero βs or three nonzero but small βs, the proposed test frequently uses one or two degrees of freedom, even as the additional β values get larger; see panels (a) and (d), (b) and (e), and (c) and (f) of Figure 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 5,000 test statistics based on data generated under the null hypothesis. As suggested in panels (a)-(c) of Figure 1, 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. Thus, the modified composite likelihood ratio test serves as an useful alternative at the price of moderate power loss.

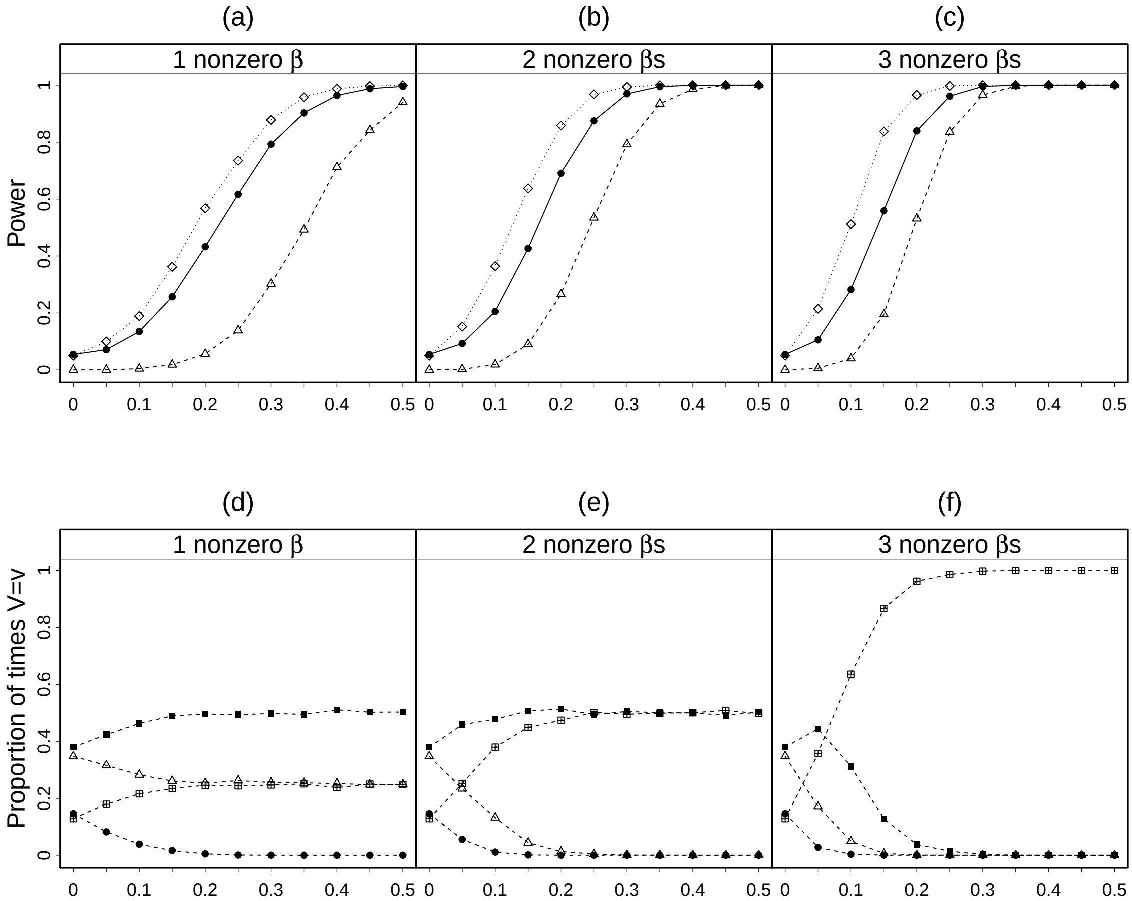

Fig. 1. Power comparisons for the matched case-control study example when one, two or three testing parameters are positive in the data generating model. (a)-(c): Power of the modified composite likelihood ratio test (●), the naive test (△), and the composite likelihood ratio test (◇) as the non-zero effect size increases from 0 to 0·5 and number of strata equals 50. (d)-(f): The proportion of times β coefficients were estimated as positive when $V = 0(\bullet), 1(\triangle), 2(\blacksquare)$ or $3(\boxplus)$ as the non-zero effect size increases from 0 to 0·5 and number of strata equals 50.

5. DISCUSSION

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

$$U_{pc}(\theta) = \frac{\partial \ell_{pc}(\theta)}{\partial \theta},\ H_p = \lim_{N\to\infty} -\frac{1}{N} E\left\{\frac{\partial^2 \ell_{pc}(\theta)}{\partial \theta^{\mathrm{T}} \partial \theta}\right\},\ V_p = \lim_{N\to\infty} \frac{1}{N} E\left\{\frac{\partial \ell_{pc}(\theta)}{\partial \theta}\right\}\left\{\frac{\partial \ell_{pc}(\theta)}{\partial \theta}\right\}^{\mathrm{T}}$$

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

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

We define the modified profile composite likelihood ratio statistic as $W_p = 2\{\ell_{MP}(\hat{\theta}_M) - \ell_{MP}(0, \eta_0)\}$, 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 methods require additional tuning parameters. Our proposed method is computationally simpler and is free of unknown tuning parameters.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes regularity conditions, proof of the asymptotic expansion in equation (2), proofs of result in Theorem 1 and a more general result with 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. *Journal of the Royal Statistical Society. Series B (Methodological)* , 239–281.

BESAG, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)* , 192–236.

CHANDLER, R. E. & BATE, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika* **94**, 167–183.

CHEN, Y. & LIANG, K.-Y. (2010). On the asymptotic behavior of the pseudolikelihood ratio test statistic with boundary problems. *Biometrika* **97**, 603–620.

CHEN, Y., NING, J., NING, Y., LIANG, K.-Y. & BANDEEN-ROCHE, K. (2017). On pseudolikelihood inference for semiparametric models with boundary problems. *Biometrika* **104**, 165–179.

GONG, G. & SAMANIEGO, F. J. (1981). Pseudo maximum likelihood estimation: Theory and applications. *Annals of Statistics* **9**, 861–869.

LIANG, K.-Y. (1987). Extended Mantel-Haenszel estimating procedure for multivariate logistic regression models. *Biometrics* **43**, 289–299.

LINDSAY, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics* **80**, 221–39.

MOLENBERGHS, G. & VERBEKE, G. (2005). *Models for Discrete Longitudinal Data*. Springer, New York.

PACE, L., SALVAN, A. & SARTORI, N. (2011). Adjusting composite likelihood ratio statistics. *Statistica Sinica* **21**, 129–148.

SELF, S. G. & LIANG, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**, 605–610.

SHAPIRO, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika* **72**, 133–144.

SUSKO, E. (2013). Likelihood ratio tests with boundary constraints using data-dependent degrees of freedom. *Biometrika* **100**, 1019–1023.

VARIN, C., REID, N. & FIRTH, D. (2011). An overview of composite likelihood methods. *Statistica Sinica* **21**, 5–42.

[*Received ????* 2015. *Revised ????* 2015]